Optimal Designs for Two-Arm Randomized Phase II Clinical Trials with Multiple Constraints

By

Wei Jiang

Submitted to the graduate degree program in Biostatistics
and the Graduate Faculty of the University of Kansas in partial fulfillment of the
requirements for the degree of Doctor of Philosophy.

Chairperson Matthew S. Mayo

Jonathan D. Mahnken

Jo A. Wick

Jianghua He

John A. Ferraro

Date Defended: November 04, 2014

The Dissertation Committee for Wei Jiang
certifies that this is the approved version of the following dissertation :

Optimal Designs for Two-Arm Randomized Phase II Clinical Trials with Multiple Constraints

Chairperson Matthew S. Mayo

Date approved: November 04, 2014

Abstract

Properly planned and executed clinical trials play a vital role for modern medicine. They have been considered as the most important studies for the development of novel treatments. In particular, designs for phase II clinical trials are crucial for new treatments development and assessment since they determine whether further definitive phase III trials are necessary. Numerous phase II designs have been developed in the last several decades. In practice, however, investigators usually need to spend time to create their own statistical algorithms since only a few conventional approaches were integrated into currently available statistical software. Therefore, developing ready-to-use packages or tools for phase II designs can accelerate, facilitate, and improve the process of designing phase II studies. Moreover, most existing phase II designs were developed in the context of sequential procedures where the main outcome is a response rate. Such designs may be inefficient when endpoints are not binary and cannot be observed within a short period. Furthermore, it has been claimed by many studies that phase II designs with a concurrent control have advantages of better patient comparability and less bias. In this dissertation, we focus on the non-sequential designs for two-arm randomized phase II clinical trials. Specifically, the first paper develops an R package for optimal designs proposed by Mayo et al. (2010) where the total sample sizes are optimized under pre-specified constraints on the standard errors of estimated efficacy rates in both control and experimental arms and the difference between the two rates. However, the designs developed in Mayo et al. (2010) are limited to dichotomous outcomes only. The second paper generalizes the original methods to designs suitable for two-arm randomized phase II trials with endpoints from the

exponential dispersion family. The new designs are generalized from a frequentist perspective and the total sample sizes are minimized using multiple constraints optimization based on standard errors. This extension is more broadly applicable to other types of study measures which include several classical distributions such as the normal, exponential, and gamma as special cases. Recently, the Bayesian analysis has become a popular and widely accepted approach to statistics due to its flexibility and ability of incorporating exiting information. The third paper further generalizes the two frequentist designs to the entire exponential family from a Bayesian perspective where the total sample sizes are optimized under constraints on the average length of posterior credible intervals of the group means and the difference between the group means.

Acknowledgements

Finishing this dissertation has been one of the most significant academic challenges I have ever had to overcome. I would never have been able to finish it without the help, guidance, and support from many people in so many ways.

I would like to express my deepest gratitude and appreciation to my committee chair, Dr. Matthew Mayo, for his guidance, understanding, caring, and patience during my graduate studies at the University of Kansas Medical Center. He is always there to help me develop my background in statistics and clinical trials, and gives me a lot of valuable suggestions with my research and future work.

I would like to gratefully thank the Department of Biostatistics at the University of Kansas Medical Center and the members of my doctoral committee for their support all along with my study. In particular, I would like to thank Dr. Jonathan Mahnken, Dr. Jianghua He, Dr. Jo Wick, and Dr. John Ferraro for their hard work, expertise, advice, and patience.

I would like to sincerely thank our Graduate Education Coordinator, Jackie Jorland, for helping me schedule and arrange my dissertation defense, as well as many other aspects during my graduate studies.

I would also like to thank my parents who are always supporting me and encouraging me.

Finally, I want to thank the person who supports me the most, my wife Yaxi Huang. Without you, I would not have been able to balance my research with everything else. I could not finish this dissertation without you by my side.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Phase II Clinical Trials

Any planned experiment designed to create preventive, diagnostic, or treatment options for individuals with given medical conditions can be called a clinical trial (Pocock, 2006). It combines reasoning from both clinical and statistical perspectives and can be applied in various clinical contexts in terms of treatment modality including drugs, devices, complementary and alternative medicine, and prevention (Piantadosi, 2005). Its vital feature is to generalize inferences regarding the results of a treatment from a limited sample to the general population of patients (Pocock, 2006). In the field of therapeutic drug development, clinical trials usually consist of four phases. Phase I trials primarily aim to test for drug safety, such as toxicity and acceptable drug dosage, based on only a small number of individuals. Phase II trials also include a small number of patients for which both the safety and pilot effectiveness are of interest. In the much larger phase III studies, researchers formally compare new treatments with alternatives, no therapy, or placebo. Finally, adverse effects and additional large-scale, long-term studies of morbidity and mortality are monitored after new drugs are approved by the Food and Drug Administration (FDA) for marketing in phase IV trials (Piantadosi, 2005; Pocock, 2006).

As explained above, a phase III trial usually requires a large sample size for evaluating efficacy and safety. This means that such trials are lengthy and expensive. Therefore, in practice, the phase III comparative testing will not be implemented unless positive evidence of safety and effectiveness is demonstrated in early phase trials (Piantadosi, 2005; Pocock, 2006). In addition, as

claimed by Piantadosi (2005), careful planning leads to a good study design and execution, which themselves lead to high-quality evidence about the new treatment effectiveness. Therefore, the planning, design, and quality of phase II clinical trials are important for new treatments development and assessment since they determine whether further definitive phase III trials are necessary (Pocock, 2006).

The term phase II clinical trial is used in a large range of studies. The simplest phase II trial design is the single-arm pilot study where the same treatment is assigned to all patients with neither concurrent control group nor variation in dose or treatment regimen included (Zohar et al., 2008). This type of phase II design is commonly used in the oncology area owing to reasons that the standard therapies may be unsuccessful and the placebo control is unethical to use (Biswas et al., 2007). Such a design may still be considered preferable even when the ethical problem of using control arm does not exist. This is because that, under this design, as many as possible subjects can be enrolled into the experimental arm with the given small sample size and the historical control data are often available for comparison (Biswas et al., 2007).

Adding a concurrent control arm is sometimes attractive and necessary for phase II designs. There are some reasons for this: first, situations for the historical control such as patient population and imaging technologies may be very different from those for the current study; second, when phase II trials aim to evaluate some new cytostatic biologic agents, instead of employing the traditional endpoints such as the response rate used in historical control, other alternative endpoints such as the progression-free survival are more appropriate; third, the classic single-arm phase II designs cannot separate trial effects such as patient selection, assessment schedule, treatment locations, etc. from the treatment effect on clinical outcomes, thus leading to multiple sources of bias (Mandrekar & Sargent, 2010). Therefore, some authors including Rubinstein et al. (2005), Wieand (2005), Taylor et al. (2006), Rubinstein et al. (2009), and Mandrekar & Sargent (2010) suggest the randomized phase II trial design due to its advantages of better patient comparability and less selection and outcome-trial effect bias.

The sequential designs further increase the complexity of phase II clinical trials. This type of

designs is important in early phase trials since it incorporates the ethical need into considerations by allowing trials stopping early due to futility or being unsafe (Stallard & Thall, 2001). In such designs, in addition to make a decision about whether the new treatment is promising for further investigation in a randomized phase III trial at the end of a study, a decision of whether the trial should be terminated early must also be made at each interim analysis. However, the sequential design may require a long study duration when the outcome cannot be observed in a timely fashion (e.g., survival or disease progression) (Mayo et al., 2010). Furthermore, several doses or treatment regimens (either with or without a control arm) can be included into the design of phase II trials where additional decision regarding which dose or treatment is the best must be made. Studies of this type are called the screening designs.

This dissertation focuses on optimal designs for the non-sequential two-arm randomized phase II clinical trials due to their potential advantages in the practice of medical research including greater compatibility to non-rapidly observable outcomes, better patient comparability, and less selection and outcome-trial effect bias. Specifically, in this dissertation, we first develop an R package for the optimal designs proposed in Mayo et al. (2010) applicable to single-stage two-arm randomized phase II clinical trials with binary outcomes. We then generalize the original approaches to the context of the exponential family from both frequentist and Bayesian perspectives. Therefore, section 1.2 of this chapter explains why an R package is developed for Mayo et al. (2010) designs. The differences in philosophy between frequentist and Bayesian approaches to sample size determination are discussed in section 1.3. Section 1.4 illustrates and summarizes the motivations and structure of this dissertation.

1.2 Why Develop a Package in R

Clinical trials are very important for the development of successful novel treatments. In practice, however, there is no uniform statistical method for designing and analyzing clinical studies since each trial has its own particular structure and goals (Piantadosi, 2005). This is also true in phase II trials where numerous designs were developed in the past several decades. Therefore, unless

a conventional design for which existing computer program is available for immediate use can be tailored to fit a particular trial, investigators often need to spend time to develop their own algorithms.

There are many statistical software and languages available of which SAS and R are probably the two most popular statistical programming languages. SAS is commonly used to conduct data analyses, power calculations, and sample size estimations in clinical studies. However, it provides limited procedures or macros in the phase II setting. The `PROC POWER` procedure in SAS provides computations for sample size and power based on the Fisher's exact test. Groulx et al. (2007) created SAS macros `%SIMON` and `%ADAPTIVE` for Simon's two-stage designs (Simon, 1989) and Lin and Shih's adaptive design (Lin & Shih, 2004), respectively. A different macro named `phase2` for Simon's two-stage designs (Simon, 1989) was developed by Cantor & Moffitt Cancer Center (2009). In addition, unlike some other lower level languages such as FORTRAN and C++ which require users having in-depth knowledge of computer programming, SAS is not very flexible for developing innovative methods.

R is widely used for the designs and analyses of clinical trials by researchers from academia, government agencies, and the pharmaceutical industry since its release (Chen & Peace, 2010). It is an open source software that not only permits modifications on source codes to handle various applications but also is quite powerful for new methods development. It can be compiled and run on various of systems including UNIX platforms, windows, and MAC OS. Currently, numerous R packages have been written and published for clinical trial applications among which, unfortunately, only a few are specific for phase II clinical trial designs. The R package named **clinfun** is frequently used for both design and analysis of phase II clinical trials (Chen & Peace, 2010). It contains functions for computing effect size, sample size, and power based on the Fisher's exact test, the two-stage boundary operating characteristics, Simon's two-stage optimal and minimax designs (Simon, 1989), and the exact single-stage design. It can also be used for defining a stopping rule and its corresponding operating characteristics for the repeated significance testing based on monitoring toxicity. In addition, the function of sample size determination for the group sequential

design in the context of phase III trials is included in the **clinfun** package (Chen & Peace, 2010).

Besides SAS and R, some other high level graphical user interface statistical software such as SPSS, Minitab, and JMP can also be used to conduct some standard analyses on clinical trial data. However, they provide very little support for the designs and sample size computations in the context of phase II trials.

As discussed above, statistical software specific for phase II trial designs are limited to several conventional approaches. In order to accelerate, facilitate, and improve the process of designing phase II studies, it is necessary to develop ready-to-use tools based on different types of phase II clinical trials. Mayo et al. (2010) developed optimal designs for single-stage two-arm randomized phase II clinical trials which minimize the total sample sizes by placing multiple constraints on proportion estimates in both control and experimental arms and the difference between proportion estimates using standard errors. A SAS macro code was provided to their designs and is available upon request from the authors. Because R is another commonly and widely used statistical programming language, we develop an R package for the designs proposed in Mayo et al. (2010) in this dissertation.

1.3 Frequentist and Bayesian Approaches for Sample Size Determination

Conducting phase II clinical trials requires a lot of preparation, among which sample size determination is one of the most important calculations because it not only has influence on a trial's conclusions and implications but also affects the length and budget (Fosgate, 2009; Zhang et al., 2011). In the past several decades, many different statistical methods have been developed for sample size optimization of phase II studies. They can be classified in terms of two philosophies of statistical thoughts - frequentist and Bayesian. Literature reviews for the phase II clinical trial designs from a frequentist perspective are discussed in sections 2.1 and 3.1. For references to phase II Bayesian designs, readers are referred to section 4.1. In this section, we review the statistical philosophies of both frequentist and Bayesian approaches, as well as their advantages and disadvantages associated with sample size determination.

The frequentist approach for determining sample size is most common (Chung & Schulz, 2007). It is broadly implemented in phase II trials and widely accepted by both medical journals and the regulators such as the FDA. In general, statistical methods from a frequentist perspective assume that the parameters of interest on which inferences center are unknown but not random, thus no prior probability distribution is associated with them. Therefore, the frequentist approach defines probabilities in terms of the data space and makes the inferences based on a particular experiment (Berry et al., 2010). When designing phase II trials, the frequentist approach utilizes prior information by estimating the values of unknown parameters according to literature, expert knowledge, or some pilot studies (Chung & Schulz, 2007). The sample size is usually determined in terms of the specified goal such as minimizing variance or maintaining adequate power (Maxwell et al., 2008).

One advantage of the frequentist approach is that it provides sample size determination some systematic and widely accepted operating characteristics such as the power and Type I error. These characteristics are very important to regulators such as the FDA (Berry et al., 2010). Moreover, because the frequentist approach does not require additional specification of prior densities, the statistical models and equations for sample size calculation are often straightforward and easy to understand and use. In addition, the frequentist approach usually does not require high computational cost and gives a standard route to evaluating the performance of a wide range of statistical methods.

A key limitation in principle to the frequentist approach is that it does not consider the uncertainty involved in estimating unknown parameters. When determining sample sizes for phase II trials, frequentist methods require particular values of unknown parameters which are not observable and need to be estimated according to external information. Misspecification of these values may lead to less than optimal designs and invalid inferences. Some limitations also arise when an early termination plan is necessary for a phase II trial due to the ethical or economic considerations. Although the sequential monitoring is possible in the frequentist setup, the frequentist approach is not very flexible to possible changes that could occur during a study. One reason is that the

error rates are affected by the inclusion of several interim analyses, so interpretations of statistical inferences are less apparent (Biswas et al., 2007; Berry et al., 2010). Some authors developed strategies for choosing stopping rules in order to control the type I error rate; for examples, see Pocock (1977), O'Brien & Fleming (1979), and Fleming et al. (1984).

The Bayesian approach is becoming popular lately owing to its flexibility and ability of incorporating existing information. In general, statistical methods from a Bayesian perspective assume the parameters of interest having prior distributions by which summary information for unknown parameters is provided prior to collecting data. Based on data once observed, prior probabilities are updated to posterior distributions for statistical inference. In other words, the posterior distribution combines the prior knowledge and data information of unknown parameters, so the inferences made from Bayesian methods include all information currently available (Berry et al., 2010). In addition, the Bayesian method is inherently adaptive, so it is not necessary to define a sample size in advance (Berry et al., 2010). However, Mayo & Gajewski (2004) and guidelines from FDA for the use of Bayesian statistics in medical device clinical trials suggest that a pretrial sample size should be determined according to safety and effectiveness endpoints (Food and Drug Administration, 2010). When designing phase II trials, the Bayesian approach utilizes external information by assigning prior distributions to the parameters of interest. The sample size is usually determined in terms of either the decision-theoretic criterion or the inferential-theoretic criterion (see section 4.1).

One advantage of the Bayesian approach is that it allows the direct use of probability distributions on the parameters of interest, thus taking the uncertainty associated with all unknowns into considerations. Another advantage is its flexibility in possible design modifications during a trial. As explained in section 1.1, the sequential design is sometimes important in early phase trials because termination plans may be necessary due to ethical or economic considerations (Stallard & Thall, 2001). The Bayesian approach is flexible in that stopping a trial early does not influence the credibility of Bayesian inferences and modifications can be made in terms of accumulated information at any time during a study (Berry et al., 2010). Furthermore, because the main purpose

of phase II trials is to evaluate whether a therapy or therapies are sufficiently effective and safe to warrant a larger and more expensive phase III study (Piantadosi, 2005; Pocock, 2006), designing a phase II clinical trial can be treated as a decision problem (Stallard, 1998; Biswas et al., 2007; Berry et al., 2010). The Bayesian approach is ideal for such situations since the costs and benefits involved in decision making can be assessed using the utility or loss function defined under the Bayesian decision-theoretic criterion (Berry et al., 2010).

The greatest concern to the Bayesian approach is the question of "objectivity". Before computing the sample size, a prior distribution must be selected for all unknown parameters, and the posterior distributions thus the results can be heavily influenced by the chosen priors. In practice, however, there is no uniform way to select prior densities, meaning that different researchers may suggest different prior distributions. Therefore, eliciting a prior that properly reflects experts' beliefs is very challenging and considered as the most crucial step in the Bayesian analysis (Spiegelhalter et al., 2004). Cautions should be taken when eliciting prior distributions or misleading results may be produced. Moreover, the Bayesian approach usually involves high-dimensional integrals for which closed form solutions often do not exist. In this case, the posterior densities are estimated using the simulation-based approaches such as the Markov Chain Monte Carlo (MCMC). These methods, however, produce random errors in results and often lead to high computational costs. Some limitations also arise from the flexibility of the Bayesian approach. The first one is that deviations from the original plan can harm some study conclusions (Berry et al., 2010; Mahajan & Gupta, 2010). Such flexibility may also prompt difficulties in controlling the overall type I error rate (Mahajan & Gupta, 2010). Besides, although the Bayesian approach is useful in solving decision making problem, clinical trial designs based on decision-theoretic methods are relatively uncommon in practice (Berry et al., 2010). One reason is that it is difficult to specify a good utility or loss function that satisfies the study's specific objectives (Adcock, 1988; Berry et al., 2010).

As discussed above, both frequentist and Bayesian approaches to sample size determinations have their advantages and disadvantages. In this dissertation, we first generalize the original designs proposed in Mayo et al. (2010) for binomial distributions only to the exponential dispersion

family of distributions from a frequentist perspective. Given that the frequentist approach may not be the best when the clinicians, for example, can only provide a range of values for the study parameters, we further generalize the two frequentist designs to the entire exponential family from a Bayesian perspective. Under different situations, each method has its own advantages and limitations. We do not advocate for using one approach over the other.

1.4 Current Studies

In the process of developing new treatments, statistics play an important and non-ignorable role since the success and quality of clinical trial researches are significantly affected by the statistical plan employed. In this dissertation, we focus on improving the statistical practice and design for single-stage two-arm randomized phase II trials.

Because well-designed phase II trials can improve the success rate in much more expensive phase III studies, a good deal of phase II designs were developed in the literature. In practice, it is impossible to propose an uniform design that is suitable for all types of phase II trials since each of them has its own particular structure and goals (Piantadosi, 2005). Tailoring a design to accommodate a particular trial at hand is difficult and time consuming, often requiring the use of computer algorithm (Thall, 2008). As discussed in section 1.2, there is a need for developing more ready-to-use packages or tools that can accelerate, facilitate, and improve phase II studies. SAS and R are the two most popular and widely used statistical software tools for the designs and analyses of clinical trials. Since the SAS macro has been already provided to the optimal designs proposed in Mayo et al. (2010), we develop an R package for the original designs in this dissertation.

The classic phase II trial designs only contain one arm, and a same treatment is assigned to all enrolled patients. Such designs may cause problems of bias since situations in previous studies may be different in many aspects from those in the current trial. Including a concurrent control arm into phase II clinical trials is sometimes attractive and necessary because it creates less bias and has advantages of better patient comparability (Rubinstein et al., 2005; Mandrekar & Sargent,

2010). Moreover, most existing designs for phase II studies were developed based on the sequential procedures that allow for incorporating the ethical and economy needs into considerations. These designs, however, could be lengthy and may not be appropriate for phase II trials where endpoints cannot be observed within a short period (Mayo et al., 2010). Mayo et al. (2010) developed the optimal designs for single-stage two-arm randomized phase II clinical trials where the total sample sizes are minimized using constraints optimization. Their designs overcome the obstacles described above, but are limited to the dichotomous endpoint. Many probability distributions from the exponential family are important in the practice of phase II clinical trials. For example, the normal distribution is useful when a phase II trial has continuous endpoints such as the change in tumor size; the exponential and gamma distributions are important for phase II studies with time to event outcomes such as the time to disease progression. Therefore, it is necessary to develop optimal designs more broadly applicable to phase II clinical studies with endpoints from distributions in the exponential family.

Over the last 10 to 20 years, the improvement of modern computing power and the development of the MCMC algorithm for sampling lead to the explosion of interest in Bayesian analysis (Berry et al., 2010). It offers a flexible way to integrate all exiting information into designs and analyses of phase II clinical trials. Unlike the frequentist method that assumes unknown parameters having fixed values, the Bayesian approach allows for uncertainty in all unknowns. Therefore, it would be worthy and beneficial to develop Bayesian optimal designs for single-stage two-arm randomized phase II studies.

The remainder of this dissertation is organized as follows. In chapter 2, an R package named **Sample.Size** is developed for the optimal designs proposed in Mayo et al. (2010). The application of this package is explained with two examples. In chapter 3, the original approaches specific for binary outcomes are generalized from a frequentist perspective to designs suitable for single-stage two-arm randomized phase II trials with endpoints from the exponential dispersion family. Its implementations are provided for multiple design considerations under the Poisson, negative binomial, normal, and exponential distributions. Chapter 4 further generalizes the two frequentist

designs to the entire exponential family from a Bayesian perspective where the total sample sizes are optimized in terms of the average length of posterior credible intervals. An illustrative example of method implementation is provided for the exponential distribution. Sample sizes are also calculated for the normal and Bernoulli distributions using two real-life phase II clinical trials. We conclude this dissertation with summary and future studies in chapter 5.

Chapter 2

An R Package for Sample Size Determination Using Optimal Designs with Multiple Constraints

by Wei Jiang, Jonathan D. Mahnken, and Matthew S. Mayo

Abstract

Previous literature (Mayo et al., 2010) proposed optimal designs for two-arm randomized phase II clinical trials where the total sample sizes are minimized under multiple constraints using the standard errors of the estimated event rates. The designs are applicable for trials with dichotomous endpoints, and can be applied to trials with or without fixed randomization allocation ratios. In this article, an R package named **Sample.Size** for conducting the methodologies developed by Mayo et al. (2010) is provided and explained. Two examples of package implementations are provided.
Key words: R package; Optimal design; Sample size; Standard error.

2.1 Introduction

Clinical trials usually consist of four phases, among which the analysis and the quality of the design of phase II trials are important for the development and assessment of a new treatment because the conclusions derived from phase II trials are determinative regarding whether further, definitive phase III trials are necessary (Pocock, 2006). Sample size calculation is crucial for the design of phase II clinical trials since it impacts the timelines through recruitment goals and the required budget of trials (Fosgate, 2009; Zhang et al., 2011).

Often, single-arm single-stage phase II studies can be designed to get an estimate on response

rate with a given level of precision; however, such a design may lack an effective termination plan when treatment is not sufficiently effective and have no control arm for comparison (Mayo et al., 2010). A termination plan is desired for the ethical and practical (e.g., budgetary) considerations, while including a concurrent control arm can protect against bias such as that exists in comparison with historical controls (Rubinstein et al., 2009). Many authors provided single-arm multi-stage phase II designs to avoid the disadvantage of lack of the termination plan; see Gehan (1961), Simon (1989), Chen et al. (1994), Herndon (1998), Sargent et al. (1999), and Friedman et al. (2010). But these designs may be inefficient in recruiting subjects for studies where outcomes cannot be observed until several months after treatment (Mayo et al., 2010). Moreover, some publications discussed designs with concurrent control including Jung (2008), Thall & Cheng (1999), Thall & Cheng (2001), Sun et al. (2009), Wason et al. (2012), and Wason & Mander (2012). However, all these randomized designs optimized the total sample size in the context of group-sequential procedures where the allocation ratio at each stage is one-to-one. This is undesired when the control arm, for example, is known to be expensive and inefficient. Under such situations, the unbalanced design may be preferred.

Mayo et al. (2010) developed single-stage randomized phase II designs in the context of allocation ratio optimization motivated by the research goal that of estimation within precision boundaries for both response rates (in treatment and control arms) and on the rate difference. In this paper, a new R package called **Sample.Size** is provided to calculate optimized sample sizes using the designs proposed in Mayo et al. (2010). **Sample.Size** can be obtained from CRAN at http://cran.r-project.org/web/packages/Sample.Size/. The function available in this package is called `Sample.Size`, which is presented and described in section 2.4.

Section 2.2 of this manuscript summarizes the optimal designs proposed by Mayo et al. (2010). Section 2.3 discusses the method of specifying values of upper bounds on multiple constraints from that work. Examples of the **Sample.Size** package are presented in Section 2.4. We conclude with a brief summary and considerations for future directions.

2.2 Optimal Designs with Multiple Constraints

In this section, we review the optimal designs proposed by Mayo et al. (2010) for randomized phase II trials which consist of a control and an experimental arm, denoted by C and E, respectively. If Y_C and Y_E are assumed to be the number of responders for arm C and arm E, then they are two independent binomially distributed random variables with parameters (n_C, π_C) and (n_E, π_E), respectively, where n_i and π_i correspond, respectively, to the sample size and assumed true event probability for arm i, $i \in \{C, E\}$. The assumed true efficacy difference between two arms is denoted by $\Delta = \pi_E - \pi_C$, so if the outcome represents being cured, the efficacy of arm E is superior to that of arm C if Δ is positive, whereas a negative value of Δ corresponds to the superiority of arm C compared to arm E.

2.2.1 Design constraints

It is well known that the maximum likelihood estimators of efficacies in control and experimental arms (i.e., π_C and π_E) and their difference (i.e., $\Delta = \pi_E - \pi_C$) are given by $\hat{\pi}_C = Y_C/n_C$, $\hat{\pi}_E = Y_E/n_E$ and $\hat{\Delta} = \hat{\pi}_E - \hat{\pi}_C$, respectively (Casella & Berger, 2002). The proposed designs optimize the total sample size, $N = n_C + n_E$, by assuring that the response in each arm and their difference are estimated with pre-specified precision using standard errors. More specifically, the three constraints under which the minimum total sample size is obtained are:

$$\begin{aligned} SE\left(\hat{\pi}_C\right) &= \sqrt{\pi_C\left(1-\pi_C\right)/n_C} \leq \gamma_C, \\ SE\left(\hat{\pi}_E\right) &= \sqrt{\pi_E\left(1-\pi_E\right)/n_E} \leq \gamma_E, \\ SE\left(\hat{\Delta}\right) &= \sqrt{\pi_C\left(1-\pi_C\right)/n_C + \pi_E\left(1-\pi_E\right)/n_E} \leq \gamma_\Delta, \end{aligned}$$

where γ_C, γ_E and γ_Δ are pre-specified upper bounds on the standard errors listed above with $0 \leq \gamma_C, \gamma_E \leq 0.5$ and $0 \leq \gamma_\Delta \leq \sqrt{0.5}$. These three inequalities can be expressed in terms of n_C and n_E

as:

$$n_C \geq \pi_C (1-\pi_C) / \gamma_C^2, \tag{2.1}$$

$$n_E \geq \pi_E (1-\pi_E) / \gamma_E^2, \tag{2.2}$$

$$n_E \geq \pi_E (1-\pi_E) \left\{ \gamma_\Delta^2 - \pi_C (1-\pi_C) n_C^{-1} \right\}^{-1}. \tag{2.3}$$

It has been shown by Mayo et al. (2010) that the equation of total sample size, $N = n_C + n_E$, is minimized under constraints (2.1) and (2.2) when $n'_C = \pi_C (1-\pi_C) / \gamma_C^2$ and $n'_E = \pi_E (1-\pi_E) / \gamma_E^2$. In order to minimize the equation of $N = n_C + n_E$ under constraint (2.3), n_E is first replaced by $f(n_C) = \pi_E (1-\pi_E) \left\{ \gamma_\Delta^2 - \pi_C (1-\pi_C) n_C^{-1} \right\}^{-1}$, and then the standard strategy that taking the derivative of N with respect to n_C and setting the result equal to zero is implemented. This leads to the conclusion that the point minimizes N under constraint (2.3) is

$$(n_C^*, n_E^*) = \left(\left[\pi_C (1-\pi_C) + \sqrt{\pi_C (1-\pi_C) \pi_E (1-\pi_E)} \right] / \gamma_\Delta^2, f(n_C^*) \right).$$

The optimal designs proposed by Mayo et al. (2010) can be applied to trials with or without fixed randomization allocation ratios, which are discussed in the next two sections.

2.2.2 Fixed ratio design

For the case of fixed allocation ratio, the optimized total sample size must satisfy both constraints (2.1) - (2.3) and the randomization relationship $n_C = r n_E$, where $r = c/e$ is the pre-specified fixed allocation ratio. Notice that both c and e are assumed to be positive integers with the greatest common divisor being 1 (i.e., relative prime). It can be shown that a two-dimensional point with positive integer coordinates whose ratio satisfies $r = c/e$ belong to the set $\{(cd, ed); d = 1, ..., \infty\}$ (Mayo et al., 2010). Therefore, with known c and e, the point with optimized sample size is derived by finding the smallest integer value of d that satisfies constraints (2.1) - (2.3). Such d can be obtained by first substituting cd and ed for n_C and n_E in inequalities (2.1) - (2.3), and then

solving each for d and rounding the largest solution up to the nearest integer. More formally,

$$d_{max} = Ceiling\left[max\left\{\frac{\pi_C(1-\pi_C)}{\gamma_C^2 c}, \frac{\pi_E(1-\pi_E)}{\gamma_E^2 e}, \frac{\pi_C(1-\pi_C)+(c/e)\pi_E(1-\pi_E)}{\gamma_\Delta^2 c}\right\}\right] \tag{2.4}$$

is the smallest integer value of d that satisfies constraints (2.1) - (2.3), and hence the desired point with the smallest sample size that satisfies both the allocation ratio and the three constraints is (cd_{max}, ed_{max}).

2.2.3 Optimal ratio design

For the case of optimal allocation ratio, the optimized total sample size is determined by diverse combinations of constraints (2.1) - (2.3), including the following four scenarios.

Scenario 1: $\left\{\left(n_C^* > n_C'\right) \cap \left(n_E^* > n_E'\right)\right\}$. This scenario implies that the optimal sample size is completely determined by constraint (2.3) and thus the point that minimizes $N = n_C + n_E$ under constraints (2.1) - (2.3) is $\left(n_C^*, n_E^*\right)$ (see Figure 2 in Mayo et al. (2010)). Since these two coordinates of this point are not necessarily positive integers, possible candidates for the optimal sample size are

$$\left(Ceiling\left[n_C^*\right], Ceiling\left[n_E^*\right]\right), \left(Floor\left[n_C^*\right], Ceiling\left[n_E^*\right]\right), \text{and } \left(Ceiling\left[n_C^*\right], Floor\left[n_E^*\right]\right),$$

among which the ones that satisfy all three constraints and have the smallest coordinate sum are the desired points.

Scenario 2: $\left\{\left(n_C^* \leq n_C'\right) \cap \left(n_E^* > n_E'\right)\right\}$. This scenario implies that the point $\left(n_C', n_E''\right)$ minimizes $N = n_C + n_E$ under constraints (2.1) - (2.3), where n_C' is determined by constraint (2.1) and $n_E'' = max\left\{f\left(n_C'\right), \pi_E(1-\pi_E)/\gamma_E^2\right\}$ (see Figure 3 in Mayo et al. (2010)). Due to the same reason that these two coordinates are not necessarily positive integers,

$$\left(Ceiling\left[n_C'\right], Ceiling\left[n_E''\right]\right) \text{ and } \left(Ceiling\left[n_C'\right], Floor\left[n_E''\right]\right)$$

are two possible candidates for the optimal sample size. The desired point is the latter if it satisfies constraints (2.1) - (2.3), and is the former otherwise.

Scenario 3: $\left\{\left(n_C^* > n_C'\right) \cap \left(n_E^* \leq n_E'\right)\right\}$. Under this scenario, $N = n_C + n_E$ is minimized by the point $\left(n_C'', n_E'\right)$, in which $n_C'' = max\left\{f^{-1}\left(n_E'\right), \pi_C\left(1-\pi_C\right)/\gamma_C^2\right\}$ and n_E' is determined by constraint (2.2) (see Figure 4 in Mayo et al. (2010)). Similar to scenarios above, two possible candidates for the optimal sample size are

$$\left(Ceiling\left[n_C''\right], Ceiling\left[n_E'\right]\right) \text{ and } \left(Floor\left[n_C''\right], Ceiling\left[n_E'\right]\right),$$

and the desired point is the one that satisfies all three constraints or the one has the smaller coordinate sum if both satisfy the three constraints.

Scenario 4: $\left\{\left(n_C^* \leq n_C'\right) \cap \left(n_E^* \leq n_E'\right)\right\}$. The optimal sample size obtained under the last scenario is completely determined by constraints (2.1) and (2.2) (see Figure 5 in Mayo et al. (2010)). Therefore, the desired point is

$$\left(Ceiling\left[n_C'\right], Ceiling\left[n_E'\right]\right).$$

In order to assess whether point $\left(Floor\left[n_C^*\right], Ceiling\left[n_E^*\right]\right)$ satisfies constraints (2.1) - (2.3), $Floor\left[n_C^*\right]$ and $Ceiling\left[n_E^*\right]$ are first substituted for c and e in equation (2.4) and then d_{max} is calculated. This point violates constraints (2.1) - (2.3) only if $d_{max} > 1$. The assessment of other points follows a similar approach. For more details on the methodology development, see Mayo et al. (2010).

2.3 Specification of Upper Bounds on Standard Errors

In practice, values of upper bounds on standard errors for constraints (2.1) - (2.3) should be specified carefully, because the required sample size relies directly on these values. Poor estimation is a result if the trial is inadequately sized, whereas unnecessarily small bound values lead to a longer duration and a more expensive trial. It requires communication between statisticians and

investigators when defining the precision levels of estimates. For investigators who are not familiar with statistics, instead of considering the values of standard error itself, they may prefer the expression that the parameters of interest vary within a specified radius (say 0.05). In such cases, values of upper bounds on standard errors can be specified by using the margin of error of corresponding asymptotic confidence intervals, where narrow $100(1-\alpha)\%$ confidence interval provides more certainty regarding the location of the true response than wider intervals. An asymptotic $100(1-\alpha)\%$ confidence interval for a parameter π is (Agresti, 2014)

$$\hat{\pi} \pm \text{margin of error} = \hat{\pi} \pm z_{\alpha/2} SE\left(\hat{\pi}\right),$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ quantile of standard normal distribution. Suppose one would like the $100(1-\alpha)\%$ confidence interval of a parameter π to be $\pm l$ or narrower, that is the margin of error $= z_{\alpha/2} SE\left(\hat{\pi}\right) \leq l$. This leads to the inequality that $SE\left(\hat{\pi}\right) \leq l/z_{\alpha/2}$. Therefore, the upper bound on standard error γ is of the following form:

$$\gamma = l/z_{\alpha/2}.$$

Values of upper bounds on standard errors for constraints (2.1) - (2.3) specified using the strategy explained in this section is described with examples in the next section.

2.4 Usage Examples

A new R package **Sample.Size** is provided to calculate optimal sample sizes for two-arm phase II clinical trials using the approaches proposed in Mayo et al. (2010). In order to illustrate the functionality of the main function `Sample.Size` of this package, two detailed examples are presented in this section, where sample sizes are computed using optimal, one-to-one, and one-to-two allocation ratios.

2.4.1 Example 1

Suppose that, in a two-arm phase II oncology study, the response rate for the standard of care is 0.2, and a new treatment is expected to have a response rate of 0.4. A researcher seeks to determine the sample size to estimate each individual efficacy with a standard error of no greater than 0.1 and the difference with a standard error of no greater than 0.15. This design question can be accomplished by using the function `Sample.Size`, whose usage with complete arguments is given as follows,

```
Sample.Size(pi_c, pi_e, gamma_c, gamma_e, gamma_delta,
    Allratio_c = NA, Allratio_e = NA).
```

The arguments `pi_c` and `pi_e` are the assumed response rates for control and experimental arms, respectively. The values of `gamma_c`, `gamma_e`, and `gamma_delta` represent, respectively, the upper bounds on the standard errors for constraints (2.1) - (2.3). The smaller values of `gamma_c`, `gamma_e`, and `gamma_delta`, the more precise the estimates. One may specify the values of `Allratio_c` and `Allratio_e` to define the allocation ratio of control arm to experimental arm (i.e., allocation ratio = `Allratio_c`/`Allratio_e`). Default values of `Allratio_c` and `Allratio_e` are both `NA` that represents missing values in R. If one leaves these two values blank, the output only includes required sample sizes for optimal and one-to-one allocation designs. Detailed description of these arguments can also be attained using the command `??Sample.Size`.

Once the package **Sample.Size** is installed on the local computer, for the above example, one needs to type the following:

```
> library(Sample.Size)
> Size1 <- Sample.Size(0.2, 0.4, 0.1, 0.1, 0.15, 1, 2)
```

Then the output is:

```
Specified values for parameters:
Response rates:
 control = 0.2 experiment = 0.4
Upper bounds for constriants:
```

```
 gammaC = 0.1 gammaE = 0.1 gammaDelta = 0.15


Required sample sizes:
[1] Optimal Design:
 nc = 16 ne = 24 n = 40
[2] 1 to 1 Allocation Design:
 nc = 24 ne = 24 n = 48
[3] 1 to 2 Allocation Design:
 nc = 16 ne = 32 n = 48
```

When upper bounds of standard errors on constraints (2.1) - (2.3) are 0.1, 0.1, and 0.15, respectively, the optimal ratio design would enroll 40 subjects in total from which 16 would be randomized to the standard of care and 24 would be assigned to the experimental arm, and one-to-one and one-to-two allocation ratio designs would both require a total of 48 subjects. Be aware that the output includes the computed sample sizes for the optimal and one-to-one allocation ratio designs by default. If designs with other allocation ratios are desired, one needs to specify the values of `Allratio_c` and `Allratio_e`. In our example, `Allratio_c` = 1 and `Allratio_e` = 2, thus the sample size for one-to-two allocation ratio design is calculated. Sometimes, one may want to estimate individual efficacies and their difference more precisely. In such cases, smaller upper bounds of standard errors on constraints (2.1) - (2.3) need to be used (e.g., see below).

```
> Size2 <- Sample.Size(0.2, 0.4, 0.05, 0.08, 0.125, 1, 2)


Specified values for parameters:
Response rates:
 control = 0.2 experiment = 0.4
Upper bounds for constriants:
 gammaC = 0.05 gammaE = 0.08 gammaDelta = 0.125
```

```
Required sample sizes:
[1] Optimal Design:
 nc = 64 ne = 38 n = 102
[2] 1 to 1 Allocation Design:
 nc = 64 ne = 64 n = 128
[3] 1 to 2 Allocation Design:
 nc = 64 ne = 128 n = 192
```

In this case, constraints chosen for constraints (2.1) - (2.3) are 0.05, 0.08, and 0.125, respectively. The total sample sizes required for the optimal, one-to-one, and one-to-two designs are 102, 128, and 192, respectively. As expected, more subjects are required for all three designs since the constraints for standard errors are more stringent here. For illustrative purposes, we also compute the sample sizes when the efficacy rate for the control arm is low as 0.05 and that for the treatment arm is 0.15, and constraints and allocation ratio are the same as `Size1`.

```
> Size3 <- Sample.Size(0.05, 0.15, 0.1, 0.1, 0.15, 1, 2)

Specified values for parameters:
Response rates:
 control = 0.05 experiment = 0.15
Upper bounds for constriants:
 gammaC = 0.1 gammaE = 0.1 gammaDelta = 0.15

Required sample sizes:
[1] Optimal Design:
 nc = 5 ne = 13 n = 18
[2] 1 to 1 Allocation Design:
 nc = 13 ne = 13 n = 26
[3] 1 to 2 Allocation Design:
```

```
 nc = 7 ne = 14 n = 21
```

The total sample sizes needed for the optimal, one-to-one, and one-to-two allocation ratio designs are 18, 26, and 21, respectively. As we can see, the required sample sizes under the same constraint criteria increase while response rates of both arms approach 0.5 for all three randomization schemes.

2.4.2 Example 2

We now consider a two-arm phase II oncology study where patients with esophageal cancer are treated with chemotherapy in control arm and with a new inhalable dry powder treatment in the experimental arm. The response rate for the standard of care is 0.35, and a new treatment is expected to have a response rate of 0.5. Researchers want to estimate each individual efficacy with a margin of error no greater than 0.15 and the difference with a margin of error no greater than 0.2.

Using the formula developed in Section 2.3 with $\alpha = 0.05$, the upper bounds for constraint (2.1) and (2.2) are both $0.15/1.96 = 0.077$, and the upper bound for constraint (2.3) is $0.2/1.96 = 0.102$. Therefore, we type in the following:

```
> Size4 <- Sample.Size(0.35, 0.5, 0.077, 0.077, 0.102, 1, 2)
```

Then the optimal sample sizes are:

```
Specified values for parameters:
Response rates:
 control = 0.35 experiment = 0.5
Upper bounds for constriants:
 gammaC = 0.077 gammaE = 0.077 gammaDelta = 0.102


Required sample sizes:
[1] Optimal Design:
 nc = 45 ne = 47 n = 92
```

```
[2] 1 to 1 Allocation Design:
 nc = 46 ne = 46 n = 92
[3] 1 to 2 Allocation Design:
 nc = 39 ne = 78 n = 117
```

A total of 92 subjects are needed for the optimal ratio design in which 45 and 47 would be randomized, respectively, to the standard of care and the experimental arms. In addition, one-to-one and one-to-two allocation ratio designs would require 92 and 117 subjects in total, respectively.

2.5 Summary and Future Directions

The **Sample.Size** package computes the required sample sizes using the optimal designs with multiple constraints proposed in Mayo et al. (2010). The methods are designed for two-arm randomized phase II clinical trials, and the required sample size can be optimized either using fixed or optimal randomization allocation ratios. The computation time of **Sample.Size** is very fast, usually does not exceed several seconds for most of the home computers. The designs can estimate individual response rates and their difference with various levels of precision using different combinations of constraints. In general, more stringent constraint criteria lead to larger sizes of sample; and notably, flexible allocation ratio can have smaller sample size requirements in various settings.

This package can also be used to design studies when only one or two of these constraints are needed. For example, if one wants to compute the sample size required for estimating the difference of response rates with pre-specified level of precision, we can define the upper bound of constraint (2.3) using the value that meets the desired precision, and using large upper boundaries of other two constraints (2.1) and (2.2). Thus, the resulting sample sizes are only affected by the constraint placing on the difference of response rates. A similar process can be followed when computing the sample sizes in the context of employing constraint(s) on an individual arm only, or on both arms but not on the difference.

One limitation to the designs proposed in Mayo et al. (2010) is that their methods are only applicable to dichotomous endpoints. We are working on generalizing the original approaches from being limited to binary endpoints to allow for outcomes from the exponential family. This future extension will be more broadly applicable to other types of study measures which include several classical distributions, such as the normal, exponential, and Poisson as special cases.

Chapter 3

Generalized Optimal Designs for Two-Arm Randomized Phase II Clinical Trials with Endpoints from the Exponential Dispersion family

by Wei Jiang, Jonathan D. Mahnken, Jianghua He, and Matthew S. Mayo

Abstract

For two-arm randomized phase II clinical trials, previous literature proposed optimal designs that minimize the total sample sizes under multiple constraints on the precision of the estimated event rates and their differences. The original designs are limited to trials with dichotomous endpoints. This paper extends the original methods to phase II clinical trials with endpoints from the exponential dispersion family distributions. The proposed optimal designs minimize the total sample sizes needed to provide estimates of population means of both arms and their difference with pre-specified precision. Sample sizes are calculated for multiple design considerations for outcomes from specific distribution families.

Key words: Multiple constraints; Optimized design; Sample size; Standard error.

3.1 Introduction

Phase II clinical trials assess both the safety and pilot effectiveness of new therapies (Pocock, 2006). It traditionally recruits a small number of patients only to the experimental arm, and the evaluated drug efficacy and safety are compared with those of a historical control (Jung, 2013). However, such traditional single-arm phase II designs may not be appropriate since patient populations in previous studies may be different in many aspects from the population in the current

trial. In addition, there are situations where very limited historical data for an existing standard therapy are available, which may be particularly true when the disease progression instead of tumor response is used as the outcome for phase II trials (Rubinstein et al., 2005). Because of these issues, Mandrekar & Sargent (2010) claimed that comparing a new study with a historical control could be questionable, and the randomized phase II design is recommended.

Multiple methods have been proposed for optimal designs for two-arm randomized phase II clinical trials. Thall & Cheng (2001) developed an optimal two-stage design based on safety and efficacy using a class of tests proposed by Thall & Cheng (1999) to minimize either the mean sample size or the maximum sample size with constraints on type I and type II errors. Jung (2008) describes a two-stage design based on testing whether the experimental arm has a higher response rate than the control arm. Sun et al. (2009) introduced a two-stage two-arm randomized design for oncology trials based on both tumor responses and early progression rates with the sample size determined by a generalization of Simon's (Simon, 1989) optimal and minimax design. Wason & Mander (2012) proposed a δ-minimax design for two-stage two-arm trails with continuous endpoints that minimizes the maximum expected sample size. This design was extended by Wason et al. (2012) to trials with more than two stages.

The term optimal for these publications is used in the context of group-sequential procedures where a one-to-one allocation ratio is performed at each stage. Multi-stage designs are sometimes inefficient in enrolling subjects especially when outcomes will not occur until several months after treatment. Furthermore, unbalanced designs often may be beneficial, especially when unequal distribution of either research or treatment costs exists between study groups. Some references in the literature discussed optimization in the context of allocation ratio; see Walter (1977), Brittain & Schlesselman (1982), Fazal (1983), Blackwelder (1986), Brooks (1987), Farrington & Manning (1990), and Sahai & Khurshid (1996). However, none of these publications present designs specifically for phase II randomized trials for which only a limited number of patients are available. Mayo et al. (2010) proposed single-stage optimal designs for two-arm randomized phase II clinical trials with dichotomous endpoints to minimize the total sample size with respect to the allocation ratio.

Their designs control precision levels of the estimated proportions in each arm and the difference between the estimated proportions. Specifically, the total sample sizes are minimized by giving individual upper bounds to the standard errors of efficacy rates of the control and experimental arms, as well as to the standard error of the difference of efficacy rates between arms.

Many probability distributions from the exponential family, such as the Bernoulli, normal, Poisson, negative binomial, exponential, and gamma distributions are important in the practice of phase II clinical trials. The binary response, defined as whether or not a patient responds to treatment, is one commonly used dichotomous outcome in phase II studies. Quite often, such a binary outcome is dichotomized from an underlying continuous response. For example, the RECIST criterion measured by a function of the change in tumor size is used for classifying a cancer patient's response to treatment into a binary response (Eisenhauer et al., 2009). However, such dichotomizing loses information (Farewell et al., 2004; Karrison et al., 2007). The normal distribution is useful in phase II clinical trials when the underlying continuous response is analyzed directly; see Karrison et al. (2007), Whitehead et al. (2009), and Wason & Mander (2012) for examples. Responses in many phase II clinical trials are in the form of counts, for example, magnetic resonance imaging lesion counts and clinical relapse counts in multiple sclerosis patients (Verhey et al., 2013), person-time response rate in breast cancer patients (O'Brien et al., 1996), and peritonitis incidence in peritoneal dialysis patients (Gadola et al., 2013). For these cases, the Poisson distribution or the negative binomial distribution can be employed. Another type of responses commonly encountered in phase II clinical trials is the time-to-event outcome, for which some examples can be found in Ameri et al. (1993), Sanson et al. (1996), Ameri et al. (1997), and Burtness et al. (2014). In this case, the exponential distribution or the gamma distribution can be used. Therefore, it is worthy to develop optimal designs for phase II clinical trials with endpoints from distributions in the exponential family. In this paper, we extend the work of Mayo et al. (2010) for dichotomous endpoints to the exponential dispersion family.

Section 3.2 of this paper describes the generalized optimal designs that place multiple constraints on the precision levels of mean estimates of both arms and their difference. This section

is divided into two parts including fixed ratio design and optimal ratio design. In section 3.3, the implementations of the proposed optimal designs are provided for the Poisson, negative binomial, normal, and exponential distributions. The total sample sizes are optimized under the optimal ratio design, as well as under one-to-two, two-to-one, and one-to-one randomization for all examples. Associated advantages and limitations of the generalized designs are discussed in section 3.4.

3.2 Generalized Optimal Designs

Throughout this paper, we denote the dispersion parameter, sample size, estimate of mean, and the standard error of the estimate as ϕ, n, $\hat{\mu}$, and $SE(\hat{\mu})$, respectively. More explicitly, ϕ_C, n_C, $\hat{\mu}_C$, and $SE(\hat{\mu}_C)$ are used for the control arm, and ϕ_E, n_E, $\hat{\mu}_E$, and $SE(\hat{\mu}_E)$ are used for the experimental arm. In addition, the difference between mean estimates and its standard error are denoted by $\hat{\Delta} = \hat{\mu}_C - \hat{\mu}_E$ and $SE(\hat{\Delta})$, respectively. The population mean is denoted by μ, thus using μ_C and μ_E for the control and experimental arms, respectively. Their difference is denoted by Δ.

3.2.1 Exponential dispersion family

Before describing the generalizations of the designs proposed by Mayo et al. (2010), some fundamental definitions and results regarding the exponential dispersion family need to be illustrated. We closely follow the discussion in Agresti (2014). Let Y be a random variable from a distribution in the exponential dispersion family, which has probability mass or density function of the form

$$f(y,\theta,\phi) = exp\left\{\frac{\theta y - \psi(\theta)}{a(\phi)} + c(y,\phi)\right\},$$

where ϕ is the dispersion parameter, and θ is the natural parameter. It can be shown that the general expressions for the mean and variance of the exponential dispersion family distributions are $E(Y) = \psi'(\theta)$ and $var(Y) = a(\phi)\psi''(\theta) = a(\phi)g(\mu)$, respectively. This indicates that the function $\psi(.)$ determines the moments of Y. The maximum likelihood estimator of the population mean $E(Y)$ is the sample mean, that is $\hat{\mu} = \mu(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} Y_i$. When $\mu(\theta)$ is an invertible function,

we have $\hat{\theta} = \mu^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$. Note that $\hat{\mu}$ is an unbiased estimator for μ, and its variance is $var(\hat{\mu}) = \frac{1}{n}a(\phi)\psi''(\theta) = \frac{1}{n}a(\phi)g(\mu)$.

3.2.2 Design constraints

This section discusses the generalized optimal designs for outcomes in the exponential dispersion family where the sample sizes are minimized by providing pre-specified precision to the mean estimates of both treatments and their difference. Specifically, the designs control the precision of the mean estimates in each arm and their difference by applying constraints on their standard errors. Therefore, our goal is to minimize the total sample size, defined by $N = n_C + n_E$, under given constraints $SE(\hat{\mu}_C) \leq \gamma_C$, $SE(\hat{\mu}_E) \leq \gamma_E$, and $SE(\hat{\Delta}) \leq \gamma_\Delta$. Note that both n_C and n_E are positive integers. With the standard errors defined previously, the total sample size N is minimized under the following constraints:

$$SE(\hat{\mu}_C) = \sqrt{\frac{a_C(\phi_C)g_C(\mu_C)}{n_C}} \leq \gamma_C$$

$$SE(\hat{\mu}_E) = \sqrt{\frac{a_E(\phi_E)g_E(\mu_E)}{n_E}} \leq \gamma_E,$$

$$SE(\hat{\Delta}) = \sqrt{\frac{a_C(\phi_C)g_C(\mu_C)}{n_C} + \frac{a_E(\phi_E)g_E(\mu_E)}{n_E}} \leq \gamma_\Delta$$

where γ_Δ, γ_C, and γ_E are all assumed to be positive. Both $a_C(\phi_C)g_C(\mu_C)$ and $a_E(\phi_E)g_E(\mu_E)$ can be zero; however, this happens only when the endpoint is a degenerate random variable which is not of interest here. Therefore, both $a_C(\phi_C)g_C(\mu_C)$ and $a_E(\phi_E)g_E(\mu_E)$ are assumed to be positive. These three constraints can be written equivalently as:

$$n_C \geq a_C(\phi_C)g_C(\mu_C)/\gamma_C^2 \tag{3.1}$$

$$n_E \geq a_E(\phi_E)g_E(\mu_E)/\gamma_E^2. \tag{3.2}$$

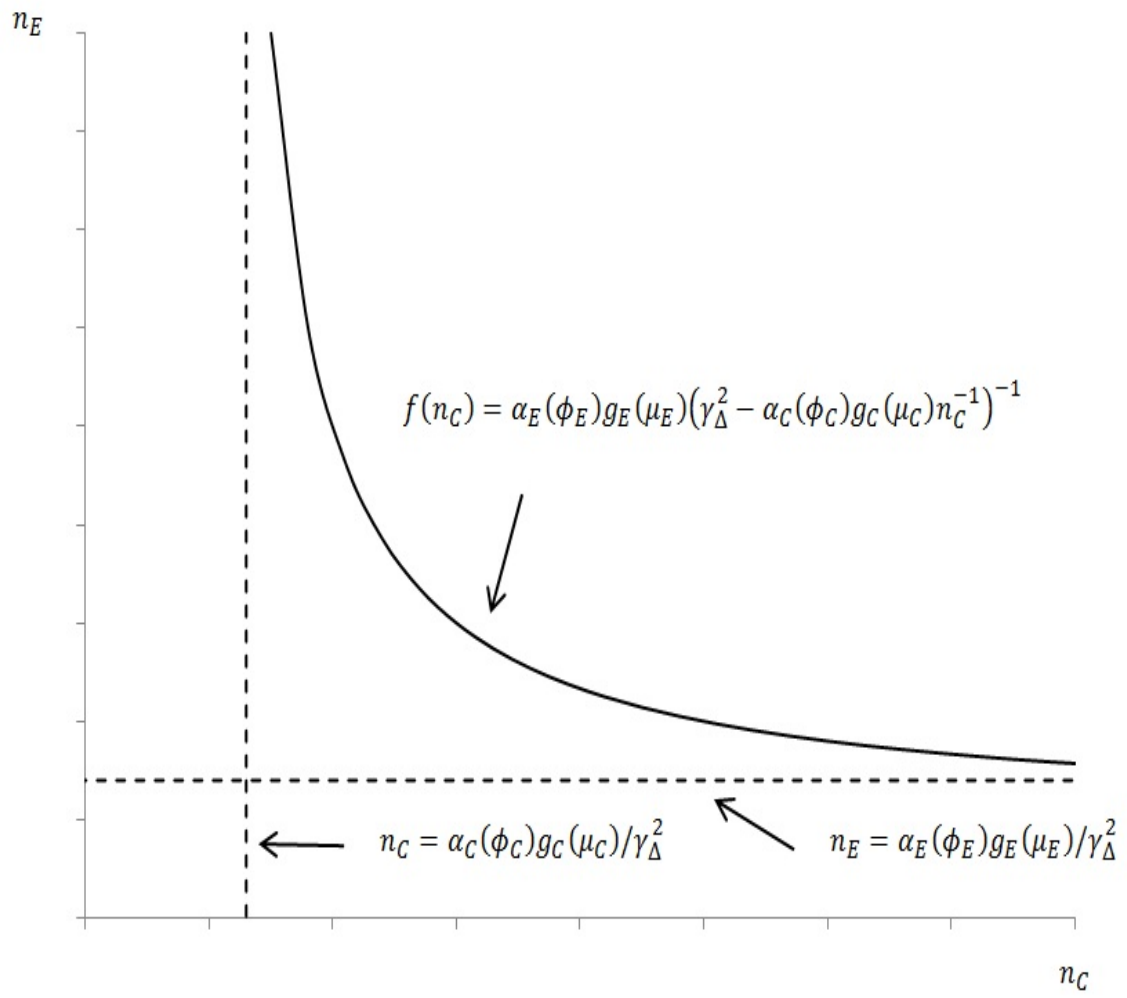

Figure 3.1: Boundary function (solid line) and its asymptotes (dashed lines).

$$n_E \geq a_E(\phi_E)g_E(\mu_E)\left\{\gamma_\Delta^2 - a_C(\phi_C)g_C(\mu_C)n_C^{-1}\right\}^{-1} \tag{3.3}$$

Because both n_C and n_E are positive integers, the boundary defined by constraint (3.3) is not functional unless it is larger than zero. Since the values of $a_E(\phi_E)g_E(\mu_E)$, γ_Δ^2, and $a_C(\phi_C)g_C(\mu_C)n_C^{-1}$ are all positive, constraint (3.3) is meaningful only if $n_C > a_C(\phi_C)g_C(\mu_C)/\gamma_\Delta^2$.

It has been shown in Mayo et al. (2010) that, for positive real numbers n_C and n_E, points that minimize $N = n_C + n_E$ are located along the bounds defined in constraints (3.1) - (3.3); therefore, the minimum value for n_C under constraint (3.1) is $n'_C = a_C(\phi_C)g_C(\mu_C)/\gamma_C^2$, and the minimum value for n_E under constraint (3.2) is $n'_E = a_E(\phi_E)g_E(\mu_E)/\gamma_E^2$. Minimum values of n_C and n_E based on

constraint (3.3) are obtained by implementing a similar strategy to that discussed in Mayo et al. (2010). Specifically, we first define the boundary function by replacing n_E and $\geq$ with $f(n_C)$ and $=$ in constraint (3.3), respectively. The resulting function is

$$f(n_C) = a_E(\phi_E)g_E(\mu_E)\left\{\gamma_\Delta^2 - a_C(\phi_C)g_C(\mu_C)n_C^{-1}\right\}^{-1},$$

where $n_C > a_C(\phi_C)g_C(\mu_C)/\gamma_\Delta^2$. The shape of this function can be determined by its several properties including that: the limit of $f(n_C)$ as n_C decreases in value approaching $a_C(\phi_C)g_C(\mu_C)/\gamma_\Delta^2$ is infinity, the limit of $f(n_C)$ when n_C goes to infinity is $a_E(\phi_E)g_E(\mu_E)/\gamma_\Delta^2$, $f'(n_C) < 0$, and $f''(n_C) > 0$. Therefore, as shown in Figure 3.1, the boundary function $f(n_C)$ defined by constraint (3.3) goes to infinity when n_C approaches $a_C(\phi_C)g_C(\mu_C)/\gamma_\Delta^2$ from the right, it decreases monotonically towards $a_E(\phi_E)g_E(\mu_E)/\gamma_\Delta^2$ as n_C approaches infinity, and it is a convex function over the range $n_C > a_C(\phi_C)g_C(\mu_C)/\gamma_\Delta^2$. Possible candidates for n_C that minimize N under constraint (3.3) are obtained by differentiating $n_C + f(n_C)$ with respect to n_C and equating the derivative to zero. The resulting equation is

$$\begin{aligned}
Q &= \frac{d}{dn_C}\left[n_C + a_E(\phi_E)g_E(\mu_E)\left\{\gamma_\Delta^2 - a_C(\phi_C)g_C(\mu_C)n_C^{-1}\right\}^{-1}\right] \\
&= 1 - a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)\left\{n_C\gamma_\Delta^2 - a_C(\phi_C)g_C(\mu_C)\right\}^{-2} \\
&= \frac{\left\{n_C\gamma_\Delta^2 - a_C(\phi_C)g_C(\mu_C)\right\}^2 - a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}{\left\{n_C\gamma_\Delta^2 - a_C(\phi_C)g_C(\mu_C)\right\}^2} \\
&= (\gamma_\Delta^4)n_C^2 - \left\{2\gamma_\Delta^2 a_C(\phi_C)g_C(\mu_C)\right\}n_C + \left\{a_C^2(\phi_C)g_C^2(\mu_C) - a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)\right\} \\
&= 0.
\end{aligned}$$

Since constraint (3.3) is defined over the range $n_C > a_C(\phi_C)g_C(\mu_C)/\gamma_\Delta^2$, the solution to the above quadratic equation is

$$n_C^* = \left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2. \tag{3.4}$$

The second derivative of $n_C + f(n_C)$ with respect to n_C verifies that n_C^* obtained from equation (3.4) is, in fact, the global minimum; therefore, $\left(n_C^*, f(n_C^*)\right)$ is the point that minimizes N under constraint (3.3). Notice that n_C', n_E', n_C^* and n_E^* may not be integers.

Similar to the designs proposed by Mayo et al. (2010), in the context of the generalized optimal designs, subjects can be randomly assigned to two arms of a phase II clinical trial either with or without a fixed allocation ratio. We call the first scheme the fixed ratio design and the second scheme the optimal ratio design, which are discussed in details in the following sections.

3.2.3 Fixed ratio design

For the fixed ratio design, the numbers of subjects assigned to the control and experimental arms are constrained by the ratio $r = c/e$, where both c and e are positive integers with the greatest common divisor being one (i.e., relative prime). This indicates that the desired point (n_C, n_E) with positive integer coordinates needs to satisfy the relationship $n_C = rn_E$ in addition to constraints (3.1) - (3.3) in such a design. It has been shown in Mayo et al. (2010) that a two-dimensional point with coordinates satisfying the ratio $r = c/e$ belong to the set $\{(cd, ed) : d = 1, 2, 3, ..., \infty\}$. Utilizing a similar method to that discussed in Mayo et al. (2010), d', the smallest integer value for d that satisfies constraints (3.1) - (3.3), is obtained by first substituting cd and ed for n_C and n_E in inequalities (3.1) - (3.3), respectively, and solving each for d, and then rounding the maximum d (say, d_{max}) up to the nearest integer. Mathematically,

$$d_{max} = max\left[\frac{a_C(\phi_C)g_C(\mu_C)}{\gamma_C^2 c}, \frac{a_E(\phi_E)g_E(\mu_E)}{\gamma_E^2 e}, \frac{\{a_E(\phi_E)g_E(\mu_E) + a_C(\phi_C)g_C(\mu_C)e/c\}}{\gamma_\Delta^2 e}\right] \tag{3.5}$$

and $d' = Ceiling(d_{max})$. With c and e known, the point of minimum sample size satisfying the allocation ratio $r = c/e$ and constraints (3.1) - (3.3) is therefore $(n_C, n_E) = (c * d', e * d')$.

3.2.4 Optimal ratio design

In many situations, the optimal ratio design may be beneficial since, compared with the fixed ratio design, it can lead to fewer subjects, thus potentially improving the economic efficiency. In this section, we develop the optimal ratio design using an analytic and a grid search methods.

3.2.4.1 Analytic method

For the optimal ratio design, the minimum sample size N is determined under various combinations of constraints (3.1) - (3.3) when there is no randomization constraint. The values of $a_C(\phi_C)g_C(\mu_C)$ and $a_E(\phi_E)g_E(\mu_E)$, and the bounds γ_C, γ_E, and γ_Δ determine under which constraint or constraints the minimum sample size is defined. As described in section 3.2.2, the minimum sample size under constraint (3.3) (i.e., $\left(n_C^*, n_E^*\right)$) is obtained by utilizing equation (3.4) and the boundary function $f(n_C^*)$; the minimum value for n_C under constraint (3.1) (i.e., n_C') is $a_C(\phi_C)g_C(\mu_C)/\gamma_C^2$; and the minimum value for n_E under constraint (3.2) (i.e., n_E') is $a_E(\phi_E)g_E(\mu_E)/\gamma_E^2$. Because n_C^* and n_E^* are either greater or not greater than n_C' and n_E', respectively, the four cases associated with the optimal ratio design are $\left\{(n_C^* > n_C') \cap (n_E^* > n_E')\right\}$, $\left\{(n_C^* \leq n_C') \cap (n_E^* > n_E')\right\}$, $\left\{(n_C^* > n_C') \cap (n_E^* \leq n_E')\right\}$, and $\left\{(n_C^* \leq n_C') \cap (n_E^* \leq n_E')\right\}$.

The first case, $\left\{(n_C^* > n_C') \cap (n_E^* > n_E')\right\}$, indicates that

$$\left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2 > a_C(\phi_C)g_C(\mu_C)/\gamma_C^2$$

and

$$f\left(\left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2\right) > a_E(\phi_E)g_E(\mu_E)/\gamma_E^2.$$

As shown in subgraph (a) of Figure 3.2, the minimum sample size is completely determined by constraint (3.3), implying that n_C^* obtained from equation (3.4) and $n_E^* = f(n_C^*)$ minimizes N. Given that $n_C^* + n_E^*$ is not necessarily an integer, $\left(Ceiling[n_C^*], Ceiling[n_E^*]\right)$ produces a point with

positive integer coordinates that satisfies constraints (3.1) - (3.3) (Mayo et al., 2010). However, this point has the potential to produce a total sample size one observation larger than the true minimum sample size under the design constraints owing to the monotone decrease of $f(n_C)$ and the rounding up to integers. Furthermore, suppose that $\left(Ceiling[n_C^*], Ceiling[n_E^*]\right)$ is an optimal point (i.e., point with the minimum sample size under the design constraints), it is possible that $\left(Ceiling[n_C^*]-1, Ceiling[n_E^*]+1\right)$ that produces the same total sample size, for example, also satisfies constraints (3.1) - (3.3). Therefore, candidates of optimal points under the first case have the form of

$$\left(Ceiling[n_C^*]-k_C-m_1, Ceiling[n_E^*]-k_E+m_1\right), (k_C,k_E)\in K, m_1\in M_1,$$

where

$$K=\{(k_C,k_E)\in\{0,1\}\times\{0,1\}; 0\le k_C+k_E\le 1\}$$

and

$$M_1=\{m_1\in\mathbb{Z}; -Ceiling[n_E^*]+k_E<m_1<Ceiling[n_C^*]-k_C, (k_C,k_E)\in K\}.$$

Here $\mathbb{Z}$ is used to denote the set of integers. Using this expression, we first let $m_1=0$ and assess whether points resulted from $(k_C,k_E)=(0,1)$ and $(1,0)$ satisfy constraints (3.1) - (3.3), and then we search values for m_1 to produce an optimal set in which every point has the minimum total sample size and satisfies the design constraints.

The second case, $\left\{(n_C^*\le n_C')\cap(n_E^*>n_E')\right\}$, indicates that

$$\left\{a_C(\phi_C)g_C(\mu_C)+\sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2\le a_C(\phi_C)g_C(\mu_C)/\gamma_C^2$$

and

$$f\left(\left\{a_C(\phi_C)g_C(\mu_C)+\sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2\right)>a_E(\phi_E)g_E(\mu_E)/\gamma_E^2.$$

In this case, subgraph (b) of Figure 3.2 suggests that $n_C'=a_C(\phi_C)g_C(\mu_C)/\gamma_C^2$ defined under con-

straint (3.1) and $n''_E = max\left\{f(n'_C), a_E(\phi_E)g_E(\mu_E)/\gamma_E^2\right\}$ with given n'_C minimize N. In general, $n'_C + n''_E$ may not be an integer; $\left(Ceiling[n'_C], Ceiling[n''_E]\right)$ could generate a total sample size one observation larger than the true minimum sample size under the design constraints; and some other points with the same total sample size may satisfy the design constraints as well. Therefore, candidates for the minimum total sample size follow the form of

$$\left(Ceiling[n'_C] - m_2, Ceiling[n''_E] - k_E + m_2\right), k_E \in \{0,1\}, m_2 \in M_2,$$

where $M_2 = \left\{m_2 \in \mathbb{Z}; -Ceiling[n''_E] + k_E < m_2 < Ceiling[n'_C], k_E \in \{0,1\}\right\}$. Using this expression, we first assess whether the point $\left(Ceiling[n'_C], Ceiling[n''_E] - 1\right)$ when $m_2 = 0$ and $k_E = 1$ satisfies constraints (3.1) - (3.3), then the set of optimal points are obtained by evaluating whether candidate points from different values of m_2 satisfy the design constraints.

The third case, $\left\{(n^*_C > n'_C) \cap (n^*_E \leq n'_E)\right\}$, indicates that

$$\left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2 > a_C(\phi_C)g_C(\mu_C)/\gamma_C^2$$

and

$$f\left(\left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2\right) \leq a_E(\phi_E)g_E(\mu_E)/\gamma_E^2.$$

As shown in subgraph (c) of Figure 3.2, in this case, $n'_E = a_E(\phi_E)g_E(\mu_E)/\gamma_E^2$ defined under constraint (3.2) and $n''_C = max\left\{f^{-1}(n'_E), a_C(\phi_C)g_C(\mu_C)/\gamma_C^2\right\}$ with n'_E given minimize N, where $f^{-1}(n'_E) = a_C(\phi_C)g_C(\mu_C)\left\{\gamma_\Delta^2 - a_E(\phi_E)g_E(\mu_E)(n'_E)^{-1}\right\}^{-1}$. Because of the same reasons stated above, the form of candidates of optimal points is defined as

$$\left(Ceiling[n''_C] - k_C - m_3, Ceiling[n'_E] + m_3\right), k_C \in \{0,1\}, m_3 \in M_3,$$

where $M_3 = \left\{m_3 \in \mathbb{Z}; -Ceiling[n'_E] < m_3 < Ceiling[n''_C] - k_C, k_C \in \{0,1\}\right\}$. The set of optimal

points are produced by first evaluating whether $\big(Ceiling[n''_C]-1, Ceiling[n'_E]\big)$ satisfies constraints (3.1) - (3.3), and then searching which values of m_3 result in the satisfaction of the design constraints.

The fourth case, $\big\{(n^*_C \le n'_C) \cap (n^*_E \le n'_E)\big\}$, indicates that

$$\left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma^2_\Delta \le a_C(\phi_C)g_C(\mu_C)/\gamma^2_C$$

and

$$f\left(\left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma^2_\Delta\right) \le a_E(\phi_E)g_E(\mu_E)/\gamma^2_E.$$

As shown in subgraph (d) in Figure 3.2, the minimum sample size is completely determined under constraints (3.1) and (3.2) in this case. This means that $n'_C = a_C(\phi_C)g_C(\mu_C)/\gamma^2_C$ and $n'_E = a_E(\phi_E)g_E(\mu_E)/\gamma^2_E$ constitute the point that minimizes the total sample size. Therefore, $(n_C, n_E) = \big(Ceiling[n'_C], Ceiling[n'_E]\big)$ is the desired point.

In this paragraph, we perform an approach similar to that discussed in Mayo et al. (2010) to assess whether a point, $\big(Ceiling[n''_C]-k_E-m_3, Ceiling[n'_E]+m_3\big)$ for example, satisfies constraints (3.1) - (3.3). Evaluations of other points can be done a similar manner. To begin with, we substitute $Ceiling[n''_C]-k_E-m_3$ and $Ceiling[n'_E]+m_3$ for c and e in $r = c/e$, respectively. This suggests that $r = (Ceiling[n''_C]-k_E-m_3)/(Ceiling[n'_E]+m_3)$ but the greatest common divisor of them is not forced to be one, which is different from the situation in the fixed ratio design. Then d_{max} is computed using equation (3.5) by following the same process as described in section 3.2.3. The point $\big(Ceiling[n''_C]-k_E-m_3, Ceiling[n'_E]+m_3\big)$ satisfies constraints (3.1) - (3.3) if d_{max} is less than or equal to one.

We have demonstrated that multiple optimal points with the same minimum total sample size may be produced for the first three cases discussed above. Mathematically, the optimal points obtained using the strategy proposed in Mayo et al. (2010) have the minimum (or the two smallest) L^p distance(s) (i.e., $||\mathbf{x}-\mathbf{y}||_p = (\sum_i |x_i - y_i|^p)^{1/p}$) to the actual points minimizing N. These points

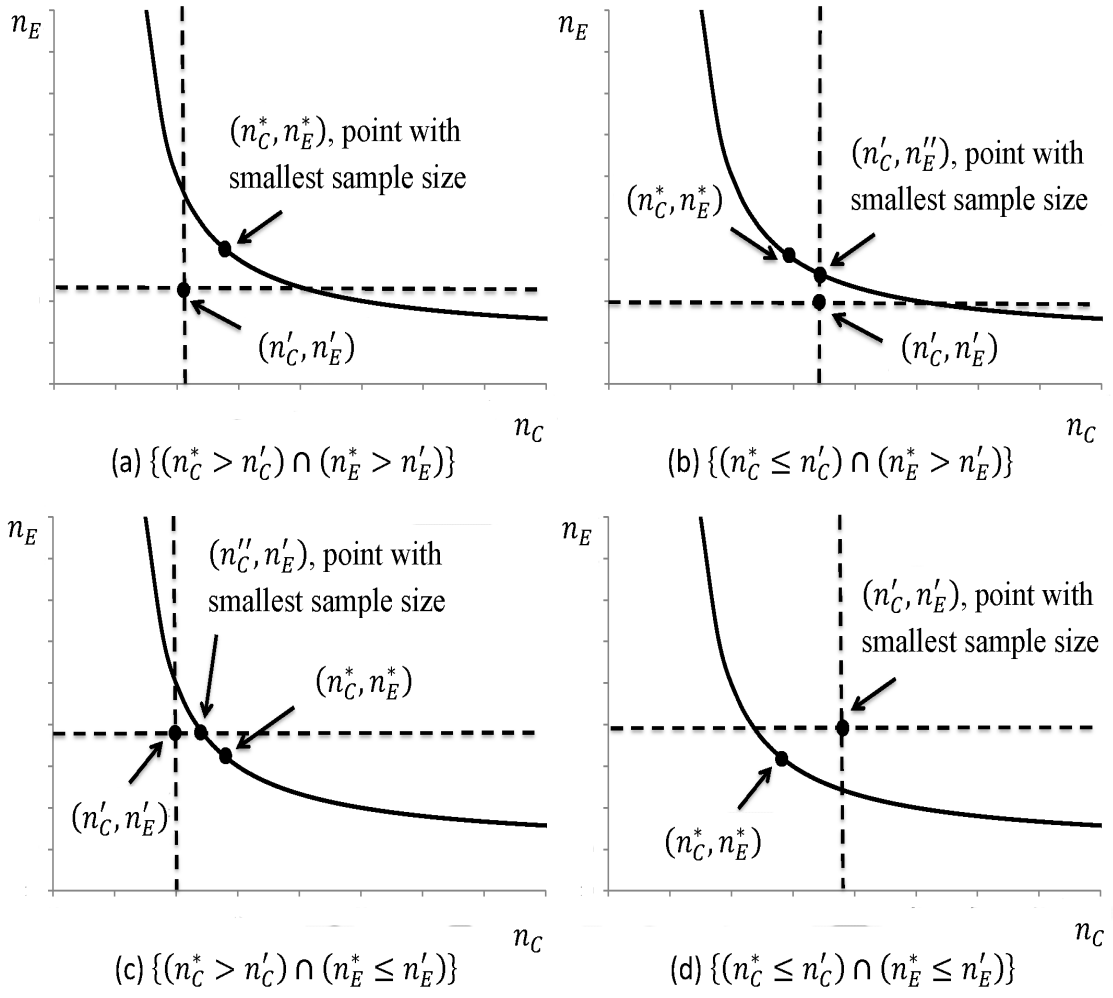

Figure 3.2: Point with the smallest sample size for four cases. Constraint (3.3) (solid line) and constraints (3.1) and (3.2) (dashed lines).

also coincide with the ones derived above with m's equal to zero. In the second case, for example, the actual point minimizing N is (n_C', n_E'') to which $\left(Ceiling[n_C'], Ceiling[n_E''] - k_E\right)$ has the shortest L_p distance compared to those with nonzero values of m_2.

3.2.4.2 Rapid grid search method

For the optimal ratio design, instead of identifying optimal points using the analytic method described previously, another stable and reliable method for finding the minimal sample size is the grid search, a numerical method utilized widely in optimization problems. The standard grid

search, however, can be extremely slow since it evaluates all combinations of sample sizes of control and experimental arms. Due to this additional computational burden and cost, in this section, we propose a rapid grid search method that can tremendously reduce the grid points searched in the sample size space and efficiently enhance the computational speed.

Figure 3.3 shows the process of the proposed rapid grid search method for the optimal ratio design. Starting from grid point $(n_C^0, n_E^0) = \big(Ceiling[n'_C], Ceiling[n'_E]\big)$, the minimum sample size determined under constraints (3.1) and (3.2), we add one sample at a time to the experimental arm until it is larger than or equal to $f(n_C^0)$ to determine an initial optimal candidate (n_C^0, n_E^1). It next moves towards right by adding one sample to n_C^0 and decreasing one sample from n_E^1 and then makes subsequent searches downwards by withdrawing one sample at a time from the experimental arm until it falls below $f(n_C^1) = f(n_C^0 + 1)$. The grid point right before n_E falls below $f(n_C^1)$ yields a second optimal candidate (n_C^1, n_E^2). After identifying k optimal candidates (say, $(n_C^{j-1}, n_E^j), j = 1, ..., k$) by repeating this procedure, we stop searching the next grid point if $n_E^k - 1$ is less than either $Ceiling[n'_E]$ or $f(n_C^{k-1} + 1)$. The desired optimal set contains those optimal candidates with the smallest total sample size. Notice that the optimal sample point is $\big(Ceiling[n'_C], Ceiling[n'_E]\big)$ if the starting value n_E^0 is larger than $f(n_C^0)$. The algorithm of the rapid grid search method is given below:

Step 1: Set the starting grid point as $(n_C, n_E) = \big(Ceiling[n'_C], Ceiling[n'_E]\big)$. If $n_E > f(n_C)$, then stop. $\big(Ceiling[n'_C], Ceiling[n'_E]\big)$ is the optimal design. If not, go to step 2.

Step 2: $n_E = n_E + 1$, and go to step 3.

Step 3: If $n_E < f(n_C)$, go back to step 2. If not, go to step 4.

Step 4: If $n_E - 1 < f(n_C + 1)$ or $n_E - 1 < Ceiling[n'_E]$, then stop and go to step 6. If not, $(n_C, n_E) = (n_C + 1, n_E - 1)$, and go to step 5.

Step 5: If $n_E - 1 > f(n_C)$, then $n_E = n_E - 1$. Keep going until $n_E - 1 < f(n_C)$, then go back to Step 4.

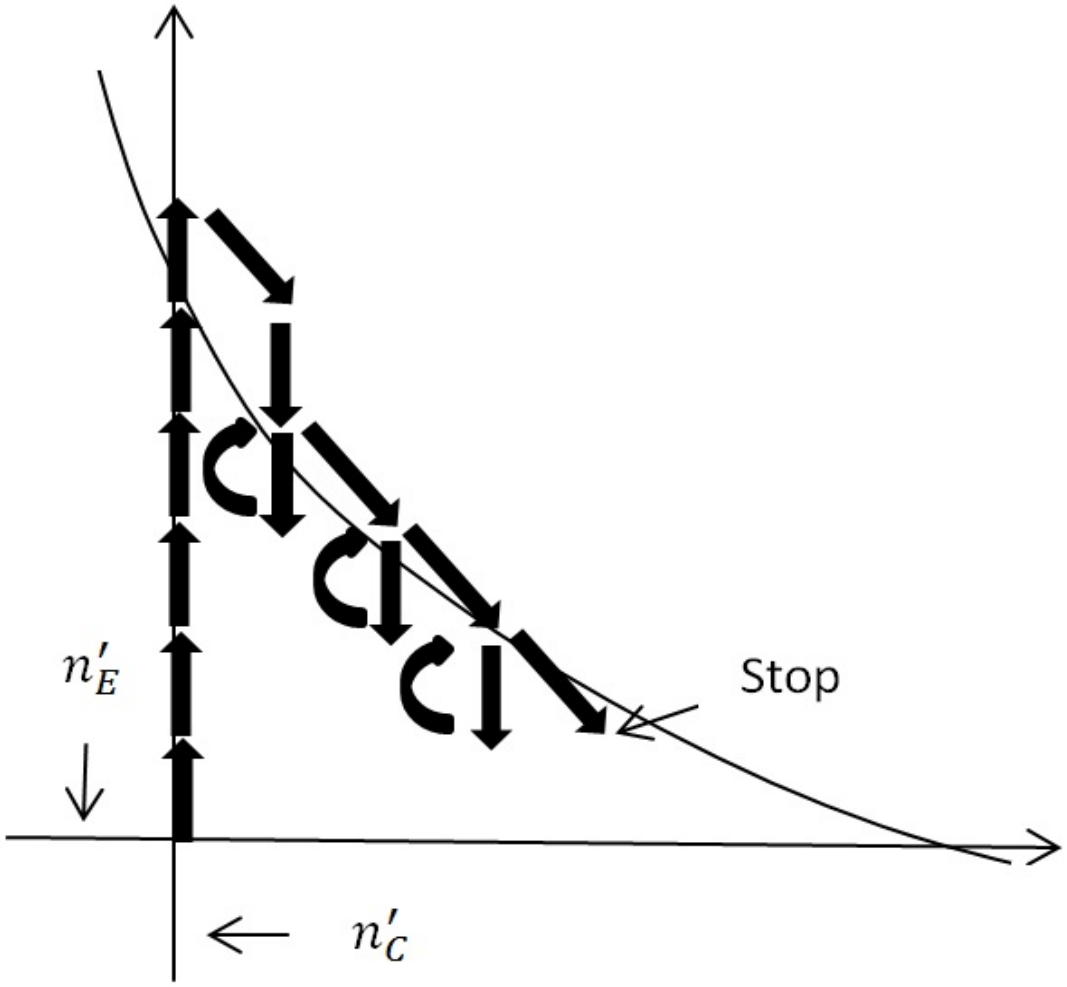

Figure 3.3: Procedure for rapid grid search method.

Step 6: Output those grid points with the smallest total sample size.

Observe that both the analytic and grid search methods may lead to multiple optimal points, and their resulting optimal sets should be identical to each other. Thus the optimal ratio design may give investigators many options, and depending on situations and objectives of phase II trials, one may be preferred over the others. For example, investigators may choose the one that assigns more subjects to the less expensive arm to reduce the trial budget. One advantage of the analytic method is that it computes the optimal point much faster than the method of rapid grid search; both methods, however, can be employed in practice.

3.2.5 Some properties of the generalized optimal designs

In previous sections, the fixed ratio and optimal ratio designs were developed. Some of their properties are discussed in this section.

Property 1. If $\gamma_E = \gamma_C$, $\gamma_\Delta^2 \geq \gamma_E^2 + \gamma_C^2 = 2\gamma_C^2$, and the variances of two arms are equal, the total sample size obtained under the optimal ratio design is generally the same as that obtained under one-to-one randomization, except that the former sometimes is 1 less than the latter.

The second case of optimal ratio design, $\left\{(n_C^* \leq n_C') \cap (n_E^* > n_E')\right\}$, is used to demonstrate this property, where n_C' under constraint (3.3) and $n_E'' = max\left\{f(n_C'), a_E(\phi_E)g_E(\mu_E)/\gamma_E^2\right\}$ are values minimizing N. When variances of two arms are equal, we have

$$\begin{aligned}
(n_C', n_E'') &= \left(a_C(\phi_C)g_C(\mu_C)/\gamma_C^2, max\left\{f(n_C'), a_E(\phi_E)g_E(\mu_E)/\gamma_E^2\right\}\right) \\
&= \left(a_C(\phi_C)g_C(\mu_C)/\gamma_C^2, max\left\{a_E(\phi_E)g_E(\mu_E)/(\gamma_\Delta^2 - \gamma_C^2), a_E(\phi_E)g_E(\mu_E)/\gamma_E^2\right\}\right) \\
&= \left(a_C(\phi_C)g_C(\mu_C)/\gamma_C^2, a_E(\phi_E)g_E(\mu_E)/\gamma_E^2\right) \\
&= \left(a_C(\phi_C)g_C(\mu_C)/\gamma_C^2, a_C(\phi_C)g_C(\mu_C)/\gamma_C^2\right)
\end{aligned}$$

for the optimal ratio design, and

$$\begin{aligned}
(n_C^\sharp, n_E^\sharp) &= \left(c * a_C(\phi_C)g_C(\mu_C)/\left(\gamma_C^2 c\right), e * a_C(\phi_C)g_C(\mu_C)/\left(\gamma_C^2 c\right)\right) \\
&= \left(a_C(\phi_C)g_C(\mu_C)/\gamma_C^2, a_C(\phi_C)g_C(\mu_C)/\gamma_C^2\right)
\end{aligned}$$

for one-to-one randomization design. Therefore, the optimal total sample size for the second case of optimal ratio deisgn is $Ceiling[n_C'] + Ceiling[n_E''] - k_E = Ceiling[n_C^\sharp] + Ceiling[n_E^\sharp] - k_E$. Since $k_E \in \{0, 1\}$, this completes the proof. It can be shown similarly that property 1 also holds for the third case of optimal ratio design. Neither assumption $\gamma_\Delta^2 \geq \gamma_E^2 + \gamma_C^2$ nor $\gamma_E = \gamma_C$ is needed in order to make this property hold for the first case of optimal ratio design, $\left\{(n_C^* > n_C') \cap (n_E^* > n_E')\right\}$, since expressions of n_C^*, n_E^*, $n_C^\sharp$, and $n_E^\sharp$ are all equal to $\left\{2a_C(\phi_C)g_C(\mu_C)\right\}/\gamma_\Delta^2$ which does not depend on γ_C and γ_E. Property 1 is valid for the fourth case of optimal ratio design, $\left\{(n_C^* \leq n_C') \cap (n_E^* \leq n_E')\right\}$,

regardless of whether $\gamma_\Delta^2 \geq \gamma_E^2 + \gamma_C^2$ or not.

Property 2. Increasing the variance of one arm with the variance of the other arm fixed, or increasing variances of both arms leads to an increase in the total sample size.

The first case of optimal ratio design, $\left\{(n_C^* > n_C') \cap (n_E^* > n_E')\right\}$, is used to demonstrate this property, where n_C^* and n_E^* are values which minimize N. It is obvious that a larger value of $n_C^* = \left\{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C)}\right\}/\gamma_\Delta^2$ is resulted from either larger values of $a_C(\phi_C)g_C(\mu_C)$ or $a_E(\phi_E)g_E(\mu_E)$ or both. The same increasing trend also holds for n_E^* since

$$\begin{aligned}
n_E^* &= a_E(\phi_E)g_E(\mu_E)\left\{\gamma_\Delta^2 - \frac{a_C(\phi_C)g_C(\mu_C)\gamma_\Delta^2}{a_C(\phi_C)g_C(\mu_C) + \sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)}}\right\}^{-1} \\
&= \left\{\frac{a_C(\phi_C)g_C(\mu_C)\gamma_\Delta^2 + \sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)}\gamma_\Delta^2 - a_C(\phi_C)g_C(\mu_C)\gamma_\Delta^2}{a_E(\phi_E)g_E(\mu_E)\left[a_C(\phi_C)g_C(\mu_C) + \sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)}\right]}\right\}^{-1} \\
&= \frac{a_E(\phi_E)g_E(\mu_E)a_C(\phi_C)g_C(\mu_C) + a_E(\phi_E)g_E(\mu_E)\sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)}}{\sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)}\gamma_\Delta^2} \\
&= \frac{\sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)} + a_E(\phi_E)g_E(\mu_E)}{\gamma_\Delta^2}.
\end{aligned}$$

It can be shown similarly that property 2 is valid for the fixed ratio design and other cases of optimal ratio design. Note that if $g_C(\mu_C)$ and $g_E(\mu_E)$ are increasing functions of μ_C and μ_E, respectively, then larger values for μ_C or μ_E or both lead to larger total sample size.

Property 3. The total sample size obtained from the optimal ratio design is always the minimum.

This property is demonstrated by comparing the sample size determined by the first case of optimal ratio design, $\left\{(n_C^* > n_C') \cap (n_E^* > n_E')\right\}$, with that determined by the fixed ratio design with one-to-one allocation ratio (i.e., $c = e = 1$). Suppose that $a_E(\phi_E)g_E(\mu_E) \geq a_C(\phi_C)g_C(\mu_C)$, then under the fixed ratio design with one-to-one allocation ratio,

$$d_{max} = max\left[\frac{a_E(\phi_E)g_E(\mu_E)}{\gamma_E^2}, \frac{a_E(\phi_E)g_E(\mu_E) + a_C(\phi_C)g_C(\mu_C)}{\gamma_\Delta^2}\right].$$

This is because the assumption $n_C^* > n_C'$ implies that $[a_C(\phi_C)g_C(\mu_C) + a_E(\phi_E)g_E(\mu_E)]/\gamma_\Delta^2$ is larger than $a_C(\phi_C)g_C(\mu_C)/\gamma_C^2$. Now denote the minimum N resulted from one-to-one fixed ratio design as $N^\sharp$, and let $N^* = n_C^* + n_E^*$. If $d_{max} = \{a_E(\phi_E)g_E(\mu_E) + a_C(\phi_C)g_C(\mu_C)\}/\gamma_\Delta^2$, then

$$\begin{aligned} N^* - N^\sharp &= \frac{-a_C(\phi_C)g_C(\mu_C) - a_E(\phi_E)g_E(\mu_E) + 2\sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)}}{\gamma_\Delta^2} \\ &= -\frac{\left[\sqrt{a_C(\phi_C)g_C(\mu_C)} - \sqrt{a_E(\phi_E)g_E(\mu_E)}\right]^2}{\gamma_\Delta^2} \leq 0. \end{aligned}$$

If $d_{max} = a_E(\phi_E)g_E(\mu_E)/\gamma_E^2$, then

$$N^* - N^\sharp \leq \frac{-a_C(\phi_C)g_C(\mu_C) - a_E(\phi_E)g_E(\mu_E) + 2\sqrt{a_C(\phi_C)g_C(\mu_C)a_E(\phi_E)g_E(\mu_E)}}{\gamma_\Delta^2}.$$

Comparing other cases of the optimal ratio design with the fixed ratio design can be shown similarly.

3.3 Implementation Examples

Section 3.2 discusses the generalized optimal designs for two-arm randomized phase II clinical trials which optimize the total sample sizes under constraints on standard errors of estimates of the population means and their difference. In this section, we first discuss strategies for specifying constraints on standard errors. The implementations of the proposed designs on the Poisson, negative binomial, normal, and exponential distributions are then given. The total sample sizes are minimized under the optimal ratio design, as well as under one-to-two, two-to-one, and one-to-one randomization ratios for each example distribution. Both the analytic and the rapid grid search methods are evaluated. Results from the analytic method are computed using macro code programmed in SAS 9.3. The rapid grid search algorithm is programmed in R 3.0.2.

3.3.1 Specification of constraints on standard errors

In practice, communications between statisticians and investigators are crucial when performing the proposed optimal designs because the resulting minimum sample sizes heavily depend on values of upper limits of constraints (3.1) - (3.3). Inefficiently large values of constraints lead to inadequately sized trials thus poor estimation, whereas lengthy and expensive trials are resulted from unnecessarily small limits. Therefore, careful evaluations are required when specifying these values. Sometimes, specifying limits on standard errors directly is not straightforward for investigators who are not familiar with statistics. If this is the case, these values can be specified using margin of errors of corresponding asymptotic confidence intervals. For example, if an investigator believes that the mean survival time of a disease should vary within a radius of 5 months, its standard error can be computed from a $100(1-\alpha)\%$ confidence interval with lower and upper limits been $\hat{\mu}-5$ and $\hat{\mu}+5$, respectively. Mathematically, an asymptotic $100(1-\alpha)\%$ Wald confidence interval for a parameter μ is defined as (Casella & Berger, 2002)

$$\hat{\mu} \pm z_{\alpha/2} SE(\hat{\mu}),$$

where $z_{\alpha/2}$ represents the $100(1-\alpha/2)\%$ quantile of standard normal distribution. Suppose that a parameter of interest is expected to have a $100(1-\alpha)\%$ confidence interval with length of $2l$ or narrower, this suggests that $SE(\hat{\mu})$ is less than or equal to $\gamma = l/z_{\alpha/2}$.

3.3.2 Poisson distribution

The Poisson distribution arises naturally when the endpoint of a phase II study takes the form of counts; for example, the person-time incidence of diseases is interested in many clinical studies. In the context of Poisson distributions, we have $a_C(\phi_C) = \phi_C = 1$, $g_C(\mu_C) = \mu_C$, $a_E(\phi_E) = \phi_E = 1$, and

$g_E(\mu_E) = \mu_E$, suggesting that the total sample size is optimized under the following constraints:

$$SE(\hat{\Delta}) = \sqrt{\frac{\mu_C}{n_C} + \frac{\mu_E}{n_E}} \leq \gamma_\Delta$$

$$SE(\hat{\mu}_C) = \sqrt{\frac{\mu_C}{n_C}} \leq \gamma_C$$

$$SE(\hat{\mu}_E) = \sqrt{\frac{\mu_E}{n_E}} \leq \gamma_E.$$

Unlike the normal distributions which have separate parameters for mean and variance, the variance of a Poisson distributed variable is equal to its expectation, implying that the difference between means affects the magnitude of the total sample size. For illustrative purposes, the optimal designs are carried out with μ_C equal to 5, 20, 30, and 40, and the value of Δ ranging from 0 to 4 by increments of 2 for $\mu_C = 5$, ranging from 0 to 8 by increments of 4 for $\mu_C = 20$, and ranging from 0 to 16 by increments of 8 for $\mu_C \geq 30$. The values chosen for γ_C and γ_E for constraints (3.1) and (3.2), respectively, are both 1.22 such that both ratios, γ_C^2/μ_C and γ_E^2/μ_E, are less than or equal to a common value 0.3. The value of γ_Δ for constraint (3.3) is set to be 1.77 and 1.48 in order to make both the 90% and 95% Wald confidence intervals for $(\mu_C - \mu_E)$ have a margin of error no greater than 2.9.

Table 3.1 lists the optimized sample sizes computed using parameter values defined above; that is, $\gamma_C = \gamma_E = 1.22$ and $\gamma_\Delta = 1.48$ and 1.77. For each set of parameters considered, minimum sample sizes are derived under the one-to-two, two-to-one, one-to-one, and optimal ratios. Note that the total sample sizes produced by one-to-two allocation ratio are identical with the ones under two-to-one allocation when two arms have equal variance. This is because that constraints (3.1) and (3.2) have the same pre-specified limits (i.e., $\gamma_C = \gamma_E$). In some cases, the minimum sample size is achieved by more than one designs. For example, when $\mu_C = 20$, $\mu_E = 28$, and $\gamma_\Delta = 1.48$, the total sample size is 44 for both optimal and one-to-one allocation ratios. For many parameter sets considered, the optimal ratio design gives multiple optimal points. When $\mu_C = 30$, $\mu_E = 46$, and $\gamma_\Delta = 1.48$, for example, the optimal set consists of $(33,36)$, $(32,37)$, $(31,38)$, $(30,39)$,

and $(29,40)$. This gives investigators more options, and which one is preferred over the others depends on situations and objectives of clinical studies. When the precision level under constraint (3.3) (i.e., γ_Δ) decreases to 1.77, the resulting total sample sizes are generally smaller than the ones obtained when $\gamma_\Delta = 1.48$. For instance, if variances of control and experimental arms are 30 and 38, respectively, the optimal ratio, one-to-two, two-to-one, and one-to-one would require enrolling, respectively, 47, 63, 78, and 52 subjects when γ_Δ is equal to 1.77. Except for the two-to-one allocation ratio, the total sample sizes of others are all smaller than the ones resulted from $\gamma_\Delta = 1.48$. The rapid grid search method is performed using $\mu_C = 30$, $\mu_E = 46$, $\gamma_C = \gamma_E = 1.22$, and $\gamma_\Delta = 1.48$, for which searching process is displayed in Figure 3.4. Identical results with the analytic method are obtained.

3.3.3 Negative Binomial distribution

The negative binomial distribution is recommended for count or frequency data with overdispersion (i.e., the variance of the outcome variable exceeds the mean), because the Poisson model may underestimate the standard errors (Berk & MacDonald, 2008). For the negative binomial distribution, we have $a_C(\phi_C) = \phi_C = 1$, $g_C(\mu_C) = \mu_C(1+k_C)$, $a_E(\phi_E) = \phi_E = 1$, and $g_E(\mu_E) = \mu_E(1+k_E)$. So the design constraints are defined as:

$$SE(\hat{\Delta}) = \sqrt{\frac{\mu_C(1+k_C)}{n_C} + \frac{\mu_E(1+k_E)}{n_E}} \leq \gamma_\Delta$$

$$SE(\hat{\mu}_C) = \sqrt{\frac{\mu_C(1+k_C)}{n_C}} \leq \gamma_C$$

$$SE(\hat{\mu}_E) = \sqrt{\frac{\mu_E(1+k_E)}{n_E}} \leq \gamma_E.$$

As is the case with the Poisson distribution, the difference between means affects the total sample size for the negative binomial distribution. For illustrative purposes, the same parameter values as those used in the Poisson example were chosen except that variances were inflated by

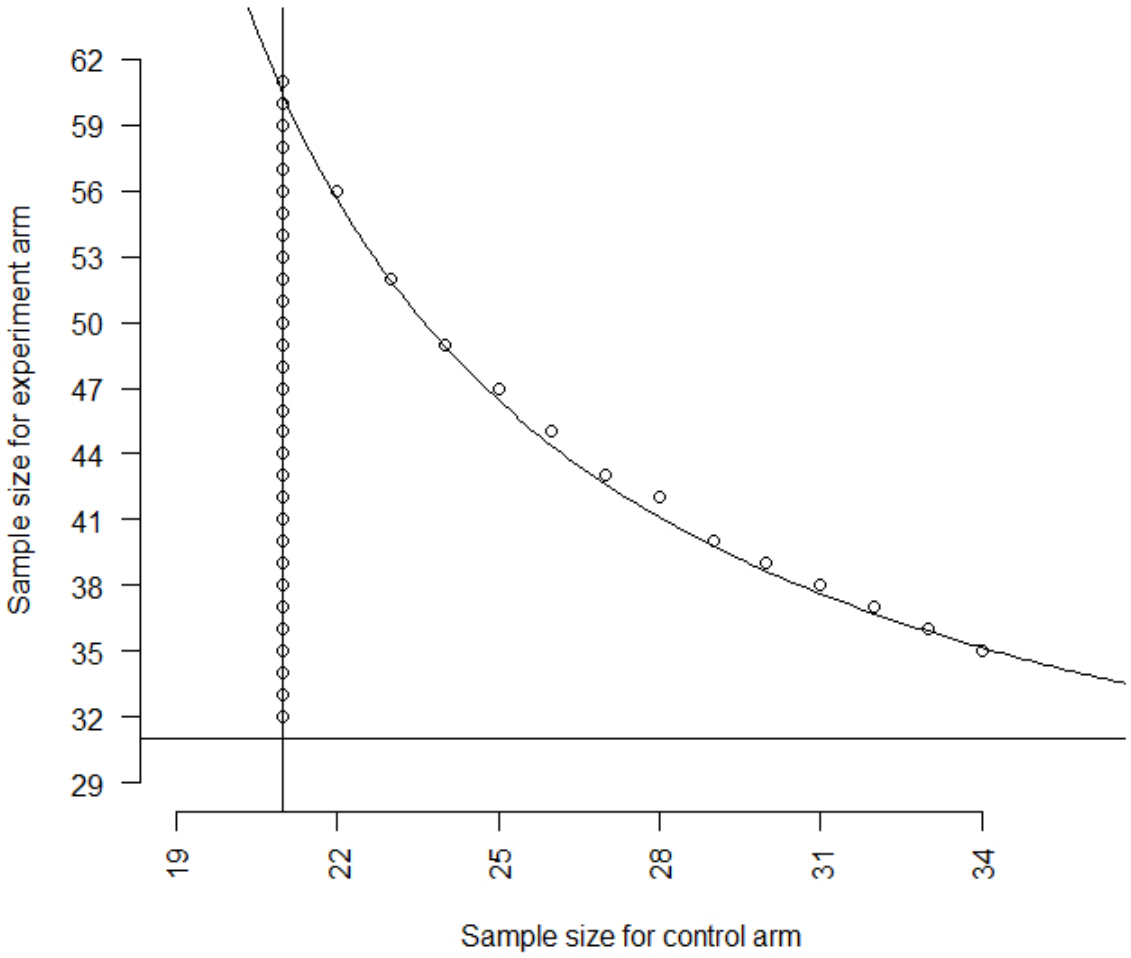

Figure 3.4: Rapid grid search procedure for Poisson distribution with $\mu_C = 30$, $\mu_E = 46$, $\gamma_C = \gamma_E = 1.22$, and $\gamma_\Delta = 1.48$. The vertical line corresponds to the minimum value determined under constraint (3.1) (i.e., n'_C). The horizontal line corresponds to the minimum value determined under constraint (3.2) (i.e., n'_E). The decreasing concave curve corresponds to the boundary function determined under constraint (3.3) (i.e., $f(n_C)$).

10% (i.e., $k_C = k_E = 0.1$) to account for overdispersion. Specifically, the values chosen for μ_C are 5, 20, 30, and 40. The values chosen for Δ are from 0 to 4 by increments of 2 when $\mu_C = 5$, from 0 to 8 by increments of 4 when $\mu_C = 20$, and from 0 to 16 by increments of 8 when $\mu_C \geq 30$.

Computed sample sizes for the negative binomial outcomes are given in Table 3.2. Observe that the results under the Poisson and the negative binomial models follow a similar pattern, except that the resulting total sample sizes from the negative binomial model are larger than those from

the Poisson model due to the inflated variances. When performing the optimal ratio design with $\gamma_\Delta = 1.48$, $\mu_C = 20$, and $\mu_E = 28$, for example, a trial would enlist 48 subjects for the negative binomial model among which 23, 22, or 21 patients and 25, 26, or 27 patients would be assigned to the control and experimental arms, respectively. In contrast, only 44 subjects would be enrolled if the Poisson model is used, where $(22,22)$, $(21,23)$, $(20,24)$, $(19,25)$, and $(18,26)$ are five different allocation schemes. In addition, Figure 3.5 shows the searching process of the rapid grid search method for the negative binomial distribution when $\mu_C = 30$, $\mu_E = 46$, $\gamma_C = \gamma_E = 1.22$, and $\gamma_\Delta = 1.48$ are employed.

Notice that the proposed designs do not require two arms having the same distribution family. When encountering the situation that an overdispersion exists in only one arm, the negative binomial distribution can be used to model the arm having overdispersion with the other remained Poisson. Suppose that in a phase II clinical study, for example, researchers believe that the most appropriate choice of means are 20 and 28 for the control and experimental arms, respectively, and the variance of the experimental arm is about 10% higher than its mean. With specified design constraints being $\gamma_C = \gamma_E = 1.22$ and $\gamma_\Delta = 1.48$, 46 subjects would be enrolled using the optimal ratio design among which 21 or 20 subjects would be randomized to the control arm and 25 or 26 subjects would be randomized to the experimental arm; on the other hand, one-to-two, two-to-one, and one-to-one allocation ratio designs would require enrolling 51, 63, and 48 subjects, respectively.

3.3.4 Normal distribution

As discussed above, the normal distribution is useful when a continuous endpoint is of interest such as the change in tumor size. When the optimal designs are performed on the normal distributions, we have $a_C(\phi_C) = \phi_C = \sigma_C^2$, $g_C(\mu_C) = 1$, $a_E(\phi_E) = \phi_E = \sigma_E^2$, and $g_E(\mu_E) = 1$, revealing that the total sample size is minimized under constraints

$$SE(\hat{\Delta}) = \sqrt{\frac{\sigma_C^2}{n_C} + \frac{\sigma_E^2}{n_E}} \leq \gamma_\Delta$$

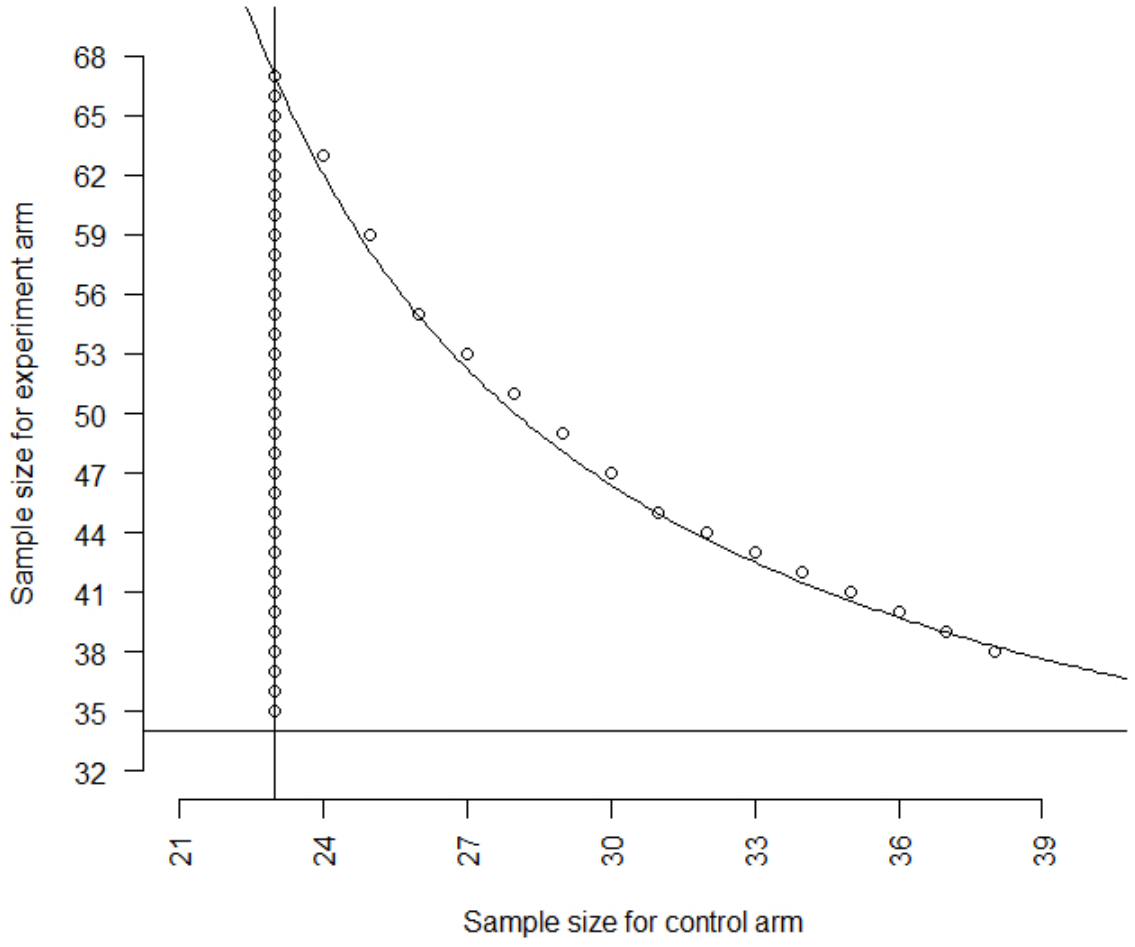

Figure 3.5: Rapid grid search procedure for negative binomial distribution with $\mu_C = 30$, $\mu_E = 46$, $\gamma_C = \gamma_E = 1.22$, and $\gamma_\Delta = 1.48$. The vertical line corresponds to the minimum value determined under constraint (3.1) (i.e., n'_C). The horizontal line corresponds to the minimum value determined under constraint (3.2) (i.e., n'_E). The decreasing concave curve corresponds to the boundary function determined under constraint (3.3) (i.e., $f(n_C)$).

$$SE(\hat{\mu}_C) = \sqrt{\frac{\sigma_C^2}{n_C}} \leq \gamma_C$$

$$SE(\hat{\mu}_E) = \sqrt{\frac{\sigma_E^2}{n_E}} \leq \gamma_E.$$

In the context of normal distribution, the difference between means is not related to the magnitude of total sample size since the variance and mean in this case are independent of each other.

For illustrative purposes, the values chosen for σ_C^2 are 5, 10, and 20, and the value of $\sigma_C^2 - \sigma_E^2$ ranges from -4 to 4 by increments of 2 when $\sigma_C^2 = 5$, and ranges from -8 to 8 by increments of 4 when $\sigma_C^2 \geq 10$. The values chosen for γ_C and γ_E for constraints (3.1) and (3.2), respectively, are both 0.77 such that both ratios, γ_C^2/μ_C and γ_E^2/μ_E, are less than or equal to a common value 0.3. The value of γ_Δ for constraint (3.3) is set to be 1.10 and 0.92 in order to make both the 90% and the 95% Wald confidence intervals for $(\mu_C - \mu_E)$ have a margin of error no greater than 1.8.

Resulting minimum sample sizes under the one-to-one, one-to-two, two-to-one, and optimal ratios when $\gamma_C = \gamma_E = 0.77$ and $\gamma_\Delta = 0.92$ and 1.10 are listed in Table 3.3. Suppose that the variances of control and experimental arms in an actual study are expected to be 20 and 24, respectively. If γ_Δ for constraint (3.3) is 1.10, then the optimal ratio design would require recruiting, respectively, 34 and 41 individuals (75 subjects in total) to the control and experimental arms; on the other hand, the one-to-two, two-to-one, and one-to-one allocation ratios would require enrolling 102, 123, and 82 subjects, respectively. The rapid grid search method for the normal distribution is performed using $\sigma_C^2 = 10$, $\sigma_E^2 = 18$, $\gamma_C = \gamma_E = 0.77$, and $\gamma_\Delta = 0.92$, for which the searching process is displayed in Figure 3.6. It produces results identical with those using the analytic method.

3.3.5 Exponential distribution

In phase II clinical trials, the exponential distribution is commonly used to model the time-to-event response and other continuous, positive, time-scaled variables. From its density function, we can see that $a_C(\phi_C) = \phi_C = 1$, $g_C(\mu_C) = \mu_C^2$, $a_E(\phi_E) = \phi_E = 1$, and $g_E(\mu_E) = \mu_E^2$. This implies that the constraints under which the total sample size is minimized are:

$$SE(\hat{\Delta}) = \sqrt{\frac{\mu_C^2}{n_C} + \frac{\mu_E^2}{n_E}} \leq \gamma_\Delta$$

$$SE(\hat{\mu}_C) = \sqrt{\frac{\mu_C^2}{n_C}} \leq \gamma_C$$

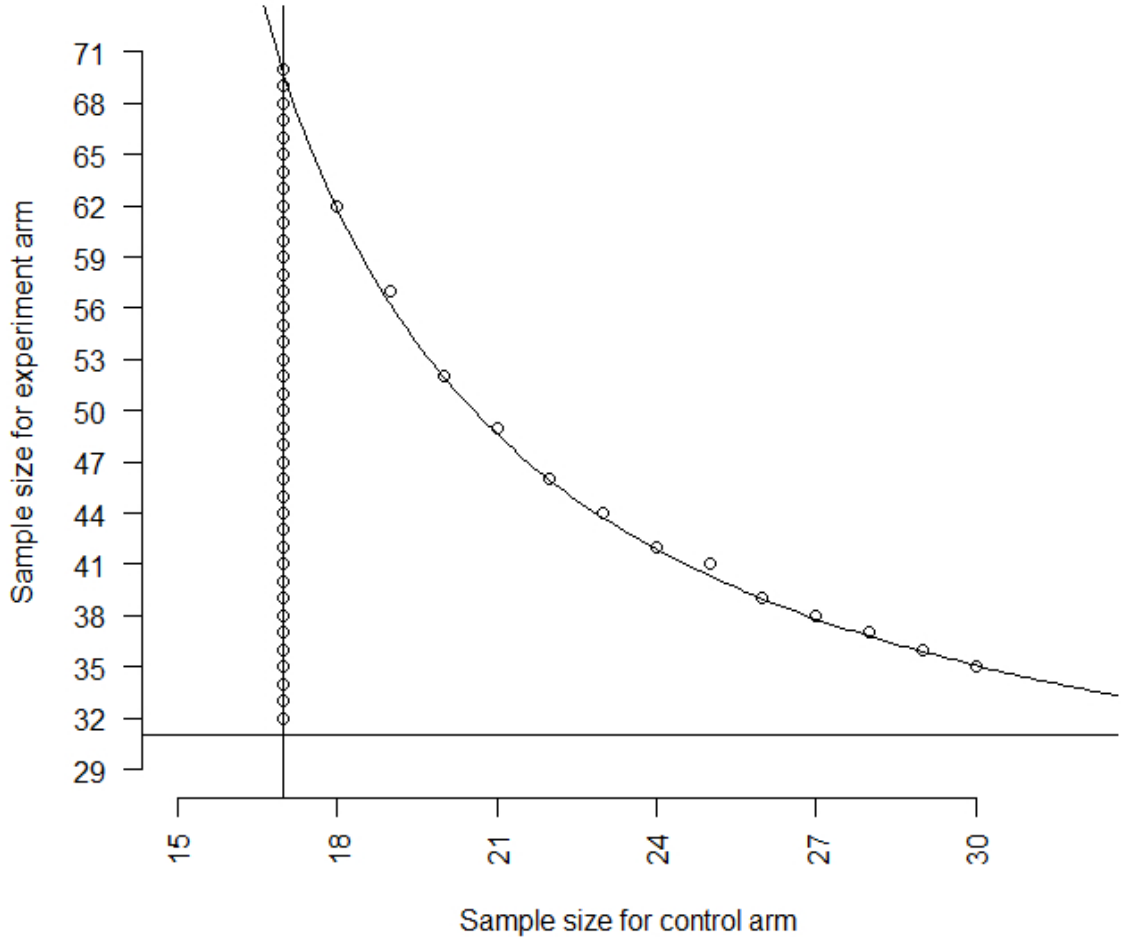

Figure 3.6: Rapid grid search procedure for normal distribution with $\sigma_C^2 = 10$, $\sigma_E^2 = 18$, $\gamma_C = \gamma_E = 0.77$, and $\gamma_\Delta = 0.92$. The vertical line corresponds to the minimum value determined under constraint (3.1) (i.e., n_C'). The horizontal line corresponds to the minimum value determined under constraint (3.2) (i.e., n_E'). The decreasing concave curve corresponds to the boundary function determined under constraint (3.3) (i.e., $f(n_C)$).

$$SE(\hat{\mu}_E) = \sqrt{\frac{\mu_E^2}{n_E}} \leq \gamma_E.$$

The variance of an exponentially distributed random variable is equal to the square of its mean, revealing that the total sample size is affected by the difference of group means. For illustrative purposes, the values chosen for μ_C are 10, 20, and 30, and the value of Δ ranges from 0 to 4 by increments of 2 when $\mu_C = 10$, and ranges from 0 to 8 by increments of 4 when $\mu_C \geq 20$. The

values chosen for γ_C and γ_E for constraints (3.1) and (3.2), respectively, are both 5.48 such that both ratios, γ_C^2/μ_C^2 and γ_E^2/μ_E^2, are not greater than a common value 0.3. The value of γ_Δ for constraint (3.3) is set to be 8.23 and 6.89 in order to make both the 90% and the 95% Wald confidence intervals for $(\mu_C - \mu_E)$ have a margin of error no greater than 13.5.

The optimal sample sizes listed in Table 3.4 are produced using $\gamma_C = \gamma_E = 5.48$, and $\gamma_\Delta = 6.89$ and 8.23 under the one-to-one, one-to-two, two-to-one, and optimal ratios. Suppose that a two-arm phase II clinical trial with an exponential outcome is now enrolling subjects. If $\gamma_\Delta = 8.23$, $\mu_C = 20$, and $\mu_E = 28$ are specified during the planning stage, under the optimal ratio design, this trial would require matriculating 41 subjects among which 14 individuals would be randomized to the control arm and 27 individuals would be randomized to the experimental arm; on the other hand, 42, 81, and 54 subjects would be enrolled, respectively, for the one-to-two, two-to-one, and one-to-one allocation designs. The rapid grid search method for the exponential distribution is conducted using $\mu_C = 30$, $\mu_E = 38$, $\gamma_C = \gamma_E = 5.48$, and $\gamma_\Delta = 6.89$, for which the searching process is displayed in Figure 3.7. It produces results identical with those using the analytic method.

3.4 Discussion

In this paper, the optimal designs for two-arm randomized phase II clinical trials with dichotomous endpoints proposed by Mayo et al. (2010) are generalized to trials with endpoints from the exponential dispersion family. The generalized optimal designs control the precision of mean estimates of both arms and their difference by applying constraints on their standard errors.

Several properties associated with the proposed designs are discussed in section 3.2.5. Property 1 points out that, under some conditions, the optimal ratio design is balanced (i.e., one-to-one) when the variances of two arms are equal; property 2 suggests that larger variances in one or both arms lead to a larger total sample size. These two properties are intuitively true since a larger sample size is required to obtain the same precision as the variances increase. Property 3 shows that the optimal ratio design always has the smallest possible sample size. This also follows our expectation because this optimal ratio is designed to minimize the total sample size.

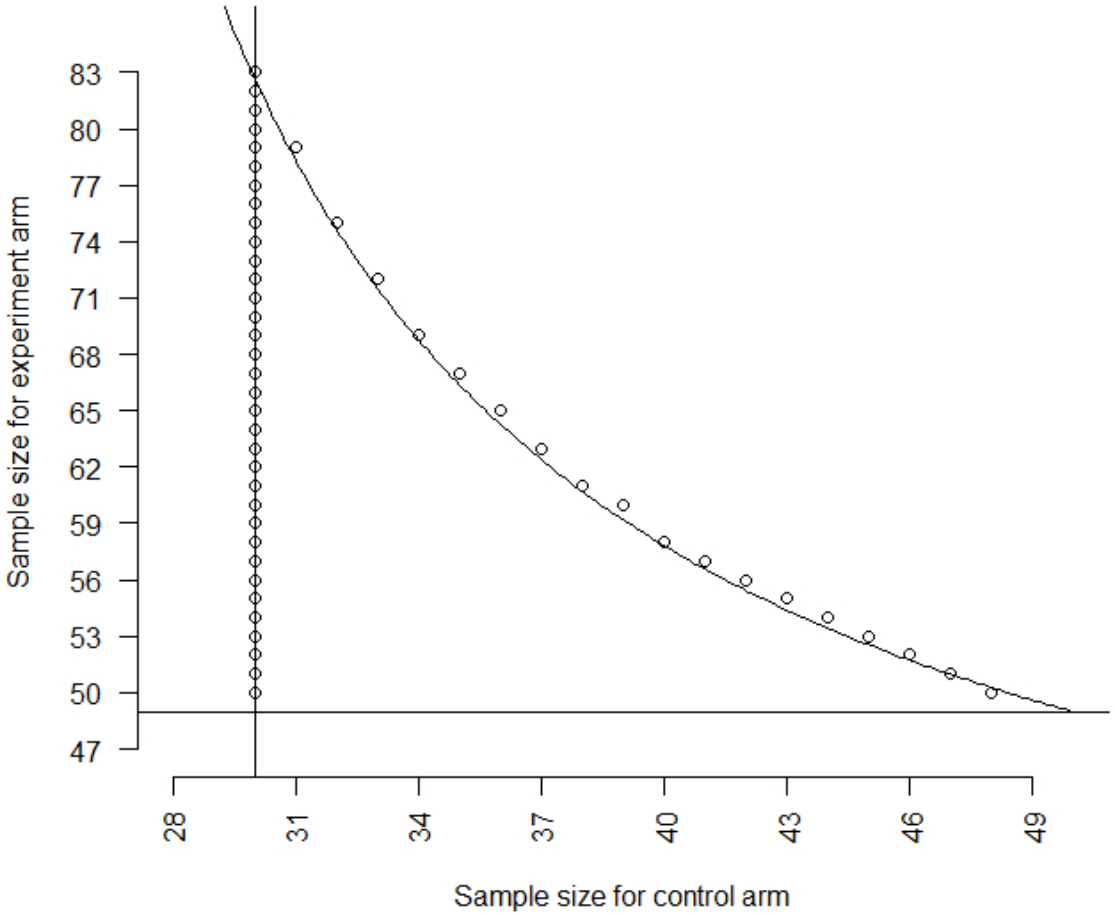

Figure 3.7: Rapid grid search procedure for exponential distribution with $\mu_C = 30$, $\mu_E = 38$, $\gamma_C = \gamma_E = 5.48$, and $\gamma_\Delta = 6.89$. The vertical line corresponds to the minimum value determined under constraint (3.1) (i.e., n'_C). The horizontal line corresponds to the minimum value determined under constraint (3.2) (i.e., n'_E). The decreasing concave curve corresponds to the boundary function determined under constraint (3.3) (i.e., $f(n_C)$).

The proposed designs use the difference between means as the measurement of associations. Notably, other approaches also exist, such as the difference between log means for Poisson and the ratio of means for the exponential distribution. The reason we use difference between means to measure the association in this article is that it is commonly used and it allows us to derive exact variances for all three constraints since often only limited sample are available in phase II settings. In the example section, the limits for constraints (3.1) - (3.3) can be obtained from the standpoint

of Wald confidence intervals. This does not mean that the proposed designs depend on the large sample approximation. Instead, compared with using the standard error directly, the expression of estimating the parameter of interest within a specified radius may be easier for investigators who are less familiar with statistics to understand. This is just an alternative way for specifying the upper limits of standard errors. In addition, both the analytic and the rapid grid search methods can be performed on the optimal ratio design; however, the analytic method allows us to compute the minimum sample size more efficiently. Sometimes, multiple optimal points are produced by the optimal ratio design. This gives more options to the investigators. Depending on different considerations, such as the cost and ethics, one may be preferred over the others.

The proposed designs are beneficial and limited in multiple aspects. In terms of advantages, first of all, the generalized approaches are very flexible in that both optimal and fixed ratio designs are allowed. For the one-to-two allocation ratio, the speed of recruitment may be accelerated by the fact that more subjects are willing to enroll in a study, especially when the standard of care is not successful (Mayo et al., 2010). Second, a variety of types of endpoints, such as the count data, dichotomous endpoints, and exponential outcomes, can be handled by the optimal designs, as long as the distributions of these endpoints belong to the exponential dispersion family. Third, the approaches do not force the outcomes in two arms belonging to the same distribution family. Therefore, some complicated situations, such as the overdispersion, can be handled by the designs. In addition, the generalized optimal designs can be easily implemented using current software such as *SAS* and *R*, so they can be readily applied when a randomized two-arm phase II clinical trial is conducted. On the other hand, one disadvantage of the proposed designs is that they are constructed based upon the exponential dispersion family. For some distributions from the more general exponential family, such as the beta distribution, the proposed methods may not work. In addition, as of all frequentist approaches, the generalized optimal designs require particular values of unknown parameters in order to compute the sample sizes. Misspecification of means and variances may lead to a less than optimal design.

Table 3.1: Minimum sample sizes when $\gamma_C = \gamma_E = 1.22$ under 1 to 2, 2 to 1, 1 to 1, and optimal randomization allocation ratios for Poisson distribution.

μ_C	μ_E	γ_Δ	1 to 2	2 to 1	1 to 1	Optimal	(n_C, n_E)
5	5	1.48	12	12	10	10	(6,4) **(5,5)** (4,6)
5	7	1.48	12	15	12	11	**(5,6)**
5	9	1.48	15	21	14	13	**(6,7) (5,8)**
20	20	1.48	42	42	38	37	(20,17) **(19,18)** **(18,19)** (17,20)
20	24	1.48	45	51	42	41	(22,19) (21,20) **(20,21)** (19,22) (18,23) (17,24)
20	28	1.48	48	57	44	44	(22,22) (21,23) **(20,24)** (19,25) (18,26)
30	30	1.48	63	63	56	55	(29,26) **(28,27)** **(27,28)** (26,29)
30	38	1.48	69	78	64	62	**(30,32) (29,33)** (28,34)
30	46	1.48	75	93	70	69	(33,36) (32,37) **(31,38)** (30,39) (29,40)
40	40	1.48	84	84	74	74	(41,33) (40,34) (39,35) (38,36) **(37,37)** (36,38) (35,39) (34,40) (33,41)
40	48	1.48	90	99	82	81	(42,39) (41,40) (40,41) **(39,42)** (38,43) (37,44) (36,45) (35,46)
40	56	1.48	96	114	88	88	(44,44) (43,45) (42,46) (41,47) **(40,48)** (39,49) (38,50) (37,51) (36,52)
5	5	1.77	12	12	8	8	**(4,4)**
5	7	1.77	12	15	10	9	**(4,5)**
5	9	1.77	12	21	14	11	**(4,7)**
20	20	1.77	42	42	28	28	**(14,14)**
20	24	1.77	42	51	34	31	**(14,17)**
20	28	1.77	42	57	38	33	**(14,19)**
30	30	1.77	63	63	42	42	**(21,21)**
30	38	1.77	63	78	52	47	**(21,26)**
30	46	1.77	63	93	62	52	**(21,31)**
40	40	1.77	81	81	54	54	**(27,27)**
40	48	1.77	81	99	66	60	**(27,33)**
40	56	1.77	81	114	76	65	**(27,38)**

Boldfaced numbers represent the optimal points having the minimum (or the two smallest) L^p distance(s) to the actual points minimizing N.

Table 3.2: Minimum sample sizes when $\gamma_C = \gamma_E = 1.22$ under 1 to 2, 2 to 1, 1 to 1, and optimal randomization allocation ratios for negative binomial distribution.

μ_C	μ_E	γ_Δ	1 to 2	2 to 1	1 to 1	Optimal	(n_C, n_E)
5	5	1.48	12	12	12	11	(7,4) **(6,5)** **(5,6)** (4,7)
5	7	1.48	15	18	14	13	(7,6) **(6,7)** (5,8)
5	9	1.48	15	21	16	14	**(6,8)**
20	20	1.48	48	48	42	41	(23,18) (22,19) **(21,20)** **(20,21)** (19,22) (18,23)
20	24	1.48	51	54	46	45	(24,21) (23,22) **(22,23)** **(21,24)** (20,25) (19,26)
20	28	1.48	54	63	50	48	(23,25) **(22,26)** (21,27)
30	30	1.48	69	69	62	61	(33,28) (32,29) **(31,30)** **(30,31)** (29,32) (28,33)
30	38	1.48	75	87	70	69	(36,33) (35,34) (34,35) **(33,36)** **(32,37)** (31,38) (30,39) (29,40)
30	46	1.48	81	102	78	76	(37,39) (36,40) (35,41) **(34,42)** (33,43) (32,44) (31,45)
40	40	1.48	93	93	82	81	(44,37) (43,38) (42,39) **(41,40)** **(40,41)** (39,42) (38,43) (37,44)
40	48	1.48	99	108	90	89	(46,43) (45,44) (44,45) **(43,46)** **(42,47)** (41,48) (40,49) (39,50)
40	56	1.48	105	126	98	96	(46,50) (45,51) **(44,52)** (43,53) (42,54)
5	5	1.77	12	12	8	8	**(4,4)**
5	7	1.77	12	18	12	10	**(4,6)**
5	9	1.77	12	21	14	11	**(4,7)**
20	20	1.77	45	45	30	30	**(15,15)**
20	24	1.77	45	54	36	33	**(15,18)**
20	28	1.77	45	63	42	36	**(15,21)**
30	30	1.77	69	69	46	46	**(23,23)**
30	38	1.77	69	87	58	52	**(23,29)**
30	46	1.77	69	102	68	57	**(23,34)**
40	40	1.77	90	90	60	60	**(30,30)**
40	48	1.77	90	108	72	66	**(30,36)**
40	56	1.77	90	126	84	72	**(30,42)**

Boldfaced numbers represent the optimal points having the minimum (or the two smallest) L^p distance(s) to the actual points minimizing N.

Table 3.3: Minimum sample sizes when $\gamma_C = \gamma_E = 0.77$ under 1 to 2, 2 to 1, 1 to 1, and optimal randomization allocation ratios for normal distribution.

σ_C^2	σ_E^2	γ_Δ	1 to 2	2 to 1	1 to 1	Optimal	(n_C, n_E)
5	1	0.92	27	15	18	13	(10,3) **(9,4)**
5	3	0.92	27	21	20	19	(12,7) **(11,8)** **(10,9)**
5	5	0.92	27	27	24	24	(13,11) **(12,12)** (11,13)
5	7	0.92	33	36	30	29	(15,14) (14,15) **(13,16)** (12,17) (11,18)
5	9	0.92	36	48	34	33	(16,17) (15,18) **(14,19)** (13,20) (12,21)
10	2	0.92	51	27	34	25	**(18,7)** **(17,8)**
10	6	0.92	51	39	38	38	(24,14) (23,15) (22,16) **(21,17)** (20,18) (19,19)
10	10	0.92	54	54	48	48	(26,22) (25,23) **(24,24)** (23,25) (22,26)
10	14	0.92	63	72	58	57	(29,28) (28,29) (27,30) **(26,31)** (25,32) (24,33)
10	18	0.92	69	93	68	65	(29,36) **(28,37)** **(27,38)** (26,39)
20	12	0.92	102	78	76	75	(45,30) (44,31) (43,32) **(42,33)** (41,34) (40,35) (39,36)
20	16	0.92	102	93	86	85	(46,39) **(45,40)** **(44,41)** (43,42)
20	20	0.92	108	108	96	95	(50,45) (49,46) **(48,47)** **(47,48)** (46,49) (45,50)
20	24	0.92	114	123	104	104	(52,52) (51,53) **(50,54)** **(49,55)** (48,56)
20	28	0.92	123	144	114	113	(55,58) (54,59) (53,60) **(52,61)** **(51,62)** (50,63) (49,64)
5	1	1.1	27	15	18	11	**(9,2)**
5	3	1.1	27	18	18	15	**(9,6)**
5	5	1.1	27	27	18	18	**(9,9)**
5	7	1.1	27	36	24	21	**(9,12)**
5	9	1.1	27	48	32	25	**(9,16)**
10	2	1.1	51	27	34	21	**(17,4)**
10	6	1.1	51	33	34	28	**(17,11)**
10	10	1.1	51	51	34	34	**(17,17)**
10	14	1.1	51	72	48	41	**(17,24)**
10	18	1.1	51	93	62	48	**(17,31)**
20	12	1.1	102	63	68	55	**(34,21)**
20	16	1.1	102	81	68	61	**(34,27)**
20	20	1.1	102	102	68	68	**(34,34)**
20	24	1.1	102	123	82	75	**(34,41)**
20	28	1.1	102	144	96	82	**(34,48)**

Boldfaced numbers represent the optimal points having the minimum (or the two smallest) L^p distance(s) to the actual points minimizing N.

Table 3.4: Minimum sample sizes when $\gamma_C = \gamma_E = 5.48$ under 1 to 2, 2 to 1, 1 to 1, and optimal randomization allocation ratios for exponential distribution.

μ_C	μ_E	γ_Δ	1 to 2	2 to 1	1 to 1	Optimal	(n_C, n_E)
10	10	6.89	12	12	10	9	**(5,4) (4,5)**
10	12	6.89	12	15	12	11	(6,5) **(5,6)** (4,7)
10	14	6.89	15	21	14	13	**(6,7) (5,8)** (4,9)
20	20	6.89	42	42	34	34	(18,16) **(17,17)** (16,18)
20	24	6.89	45	60	42	41	(20,21) **(19,22) (18,23)**
20	28	6.89	51	81	54	49	(22,27) **(21,28) (20,29)** (19,30)
30	30	6.89	90	90	76	76	(39,37) **(38,38)** (37,39)
30	34	6.89	96	117	88	87	(44,43) (43,44) (42,45) **(41,46)** (40,47) (39,48) (38,49) (37,50)
30	38	6.89	105	147	100	98	(47,51) (46,52) (45,53) (44,54) **(43,55)** (42,56) (41,57) (40,58)
10	10	8.23	12	12	8	8	**(4,4)**
10	12	8.23	12	15	10	9	**(4,5)**
10	14	8.23	12	21	14	11	**(4,7)**
20	20	8.23	42	42	28	28	**(14,14)**
20	24	8.23	42	60	40	34	**(14,20)**
20	28	8.23	42	81	54	41	**(14,27)**
30	30	8.23	90	90	60	60	**(30,30)**
30	34	8.23	90	117	78	69	**(30,39)**
30	38	8.23	90	147	98	79	**(30,49)**

Boldfaced numbers represent the optimal points having the minimum (or the two smallest) L^p distance(s) to the actual points minimizing N.

Chapter 4

Bayesian Optimal Designs for Two-Arm Randomized Phase II Clinical Trials with Endpoints from the Exponential family

by Wei Jiang, Jo A. Wick, Jianghua He, Jonathan D. Mahnken, and Matthew S. Mayo

Abstract

Frequentist optimal designs for two-arm randomized phase II clinical trials with outcomes from the exponential dispersion family were proposed by Jiang et al. (2014), where the total sample sizes are minimized under multiple constraints on the standard errors of the estimated group means. The designs were generalized from approaches developed in Mayo et al. (2010) for dichotomous outcomes. Compared to the frequentist method, the Bayesian approach offers a flexible way to incorporate uncertainty in parameters of interest into considerations. In this paper, we develop Bayesian optimal deigns for phase II clinical trials with endpoints from the exponential family from the two previous frequentist approaches. The proposed optimal designs minimize the total sample sizes under pre-specified constraints on the expected length of posterior credible intervals for both group means and their difference. The designs are applicable for trials with fixed or optimal randomization allocation ratio. Examples of method implementations are provided for different types of endpoints in the exponential family.

Key words: Multiple constraints; Optimized design; Sample size; Posterior credible interval; Monte Carlo; Natural conjugate prior family.

4.1 Introduction

Sample size calculation plays an important role in clinical research since it not only affects the success of trials but also has influence on the budget (Fosgate, 2009; Zhang et al., 2011). A recent article by Jiang et al. (2014) describes frequentist approaches to sample size determination for two-arm randomized phase II clinical trials with endpoints from the exponential dispersion family using constrained optimization. Their approaches are generalized from the ideas proposed by Mayo et al. (2010) in which the total sample sizes are optimized in the context of a binary outcome only. For both approaches, the total sample sizes are minimized by placing multiple constraints on the estimates of mean response in both control and experimental arms and the estimated difference between means using standard errors of estimates. These multiple constraints, alternatively, can be expressed in terms of confidence intervals as $L_\alpha(\mu_C) \leq l_C$, $L_\alpha(\mu_E) \leq l_E$, and $L_\alpha(\mu_E - \mu_C) \leq l_\Delta$, where $L_\alpha(.)$ represents the length of $100(1-\alpha)\%$ confidence interval for the parameter of interest, μ_C and μ_E denote the group means, and l_C, l_E, and l_Δ are cut-off lengths.

Similar to other frequentist methods of sample size determination, the two designs proposed by Jiang et al. (2014) and Mayo et al. (2010) require information regarding study parameters to compute the total sample sizes given some pre-specified constraints. In other words, the clinicians must pre-specify the values of μ_C and μ_E based on literature, expert knowledge, or some pilot studies, and misspecificaion of group means may lead to a poor estimate of the necessary sample size (Mayo et al., 2010). In practice, however, there is uncertainty regarding μ_C and μ_E since they are unknown. In many cases, for example, it is not uncommon for a clinician to only provide a range of values for the study parameter. Under such situations, it may be inappropriate to apply the above two designs since they treat the preliminary guesses as true values of parameters μ_C and μ_E, and fail to incorporate the uncertainty inherent in using the best-guesses of values into sample size optimization.

Alternatively, the Bayesian approach considers the parameters of interest as random variables, and aims to take the uncertainty in parameters into account. In the last several decades, the

Bayesian criteria of sample size computation generally consist of two classes. One is the decision-theoretic or fully Bayesian approach that maximizes the pre-selected utility function, or equivalently, minimizes the corresponding loss function for determination of sample size. The fully Bayesian approach of sample size problem was first explained in details by Raiffa & Schlaifer (1961). For other examples, see Staquet & Sylvester (1977), Sylvester (1988), Brunier & Whitehead (1994), Joseph & Wolfson (1997), Pham-Gia (1997), Lindley (1997), Stallard (1998), Pham-Gia & Turkkan (2003), and Sahu & Smith (2006). However, in the planning of a clinical trial, it may be impossible, or, at least, is time consuming to develop a reasonable utility function or loss function suitable for the study's specific objectives. Also, it is very hard to develop a widely accepted loss or utility function since different studies have different needs (Adcock, 1988). Therefore, some Bayesians recommend the performance-based or inferential-theoretic approach that controls posterior inference for the parameter of interest with certain precision. Adcock (1997) reviewed the inferential-theoretic approach without utility or loss functions. A comparison among three most commonly referenced performance-based sample size criteria, the average coverage criterion (ACC), the average length criterion (ALC), and the worst outcome criterion (WOC), which compute the sample size in terms of the length and coverage of posterior credible intervals, was provided in Cao et al. (2009). For other examples, see Adcock (1988), Joseph et al. (1995), Joseph et al. (1997), Joseph & Belisle (1997), Tan & Machin (2002), Mayo & Gajewski (2004), Clarke & Yuan (2006), Gajewski & Mayo (2006), M'Lan et al. (2008), and Hand et al. (2011).

In the literature, most Bayesian optimal designs for phase II studies are discussed in the context of adaptive designs where modifications or changes on trial and/or statistical procedure are allowed while conducting a clinical trial (Chow et al., 2008). Thall & Simon (1994) developed a Bayesian inference-based sequential design for single-arm phase II trials where the outcome is binary and the data are continuously monitored. Stallard (1998) developed a Bayesian decision-theoretic group sequential design for single-arm phase II clinical studies with a dichotomous outcome using a gain function that concentrates in the financial costs of and potential profits from the drug development program. Thall et al. (1995) presented a Bayesian single-arm phase II trial design that allows for

early termination of studies in terms of both adverse events and efficacy without explicit specification of a loss function. Stallard et al. (1999) developed a single-arm phase II sequential design that accommodates both efficacy and safety using Bayesian decision theory. Tan & Machin (2002) described sample size calculations for single-arm two-stage phase II clinical trails with a binary endpoint using the performance-based approach. Ding et al. (2008) proposed a decision-making design for single-arm phase II screening trials with binary outcomes where sampling costs and possible future payoff are incorporated using a utility function. Zhong et al. (2013) proposed a two-stage two-arm fully Bayesian design for phase II binary response trials in which the sample size can be reestimated at the interim analysis.

As claimed by Mayo & Gajewski (2004), it is necessary to compute a sample size before the trial begins for practical reasons such as budgeting. Moreover, guidelines from the Food and Drug Administration (FDA) for the use of Bayesian statistics in medical device clinical trials suggest that a minimum pretrial sample size should be determined in terms of safety and effectiveness endpoints (Food and Drug Administration, 2010). However, because more efforts regarding Bayesian designs are made in the context of adaptive procedures, the sample size problem assuming a fixed trial size are less frequently discussed in the literature. Staquet & Sylvester (1977), Sylvester (1988), and Brunier & Whitehead (1994) developed approaches of sample size calculation for single-arm single-stage phase II trials with binary responses based on the Bayesian decision theory. Mayo & Gajewski (2004) described an inferential-theoretic sample size approach for single-arm single-stage phase II studies with binary endpoints. Their method was extended from the approach developed by Tan & Machin (2002) based on diffuse prior to informative prior distributions using various strategies that allow for incorporating pessimistic and optimistic priors into sample size computation. This approach was further extended by Gajewski & Mayo (2006) to include a mixture of informative prior distributions. However, all of these designs are for dichotomous response and only contain one treatment arm. Randomized phase II trials are recommended when historical control data are limited or unreliable (Rubinstein et al., 2005; Mandrekar & Sargent, 2010). Furthermore, as demonstrated in Jiang et al. (2014), other probability distributions from the ex-

ponential family are important in the practice of phase II clinical trials. For example, Karrison et al. (2007) designed a phase II cancer trial using a continuous outcome of change in tumor size assumed to follow a normal distribution. Likewise, the exponential and gamma distributions are very useful for time-to-event outcomes such as the time-to-disease progression.

Some Bayesian sample size computations have been developed for distributions from the exponential family in the context of a two-sample comparison. The difference between two binomial proportions was considered by Joseph & Wolfson (1997) based on highest posterior density intervals. Joseph & Belisle (1997) and Pezeshk (2002) applied the Bayesian inference-based approach to determine the sample size for the differences between two normal means. Hand et al. (2011) developed an inferential-theoretic method for sample size determination in the context of comparison of two Poisson rates. However, none of these methods present designs specifically for phase II randomized trials.

In this paper, we extend the works of Mayo et al. (2010) and Jiang et al. (2014) for two-arm randomized phase II clinical trials to the entire exponential family from a Bayesian perspective. For this initial development, we assume that all hyperparameters are known. An optimal design based on hierarchical models is left for future study. Furthermore, we employ the ALC criterion introduced by Joseph et al. (1995) to determine the sample size for the following reasons: first, the ALC criterion has practical advantages in that neither utility nor loss function needs to be formulated; second, the ALC criterion is easy to understand and implement; third, the ALC criterion computes the sample size by controlling the average length of posterior credible intervals with a fixed coverage rate which is, in some sense, comparable to the frequentist approaches proposed by Mayo et al. (2010) and Jiang et al. (2014) for which design constraints can be alternatively expressed according to confidence intervals. However, we do not advocate for using one approach over the other.

The remainder of this manuscript is organized as follows. In section 4.2, we review the ALC criterion. In section 4.3, the Bayesian optimal designs that minimize the total sample sizes by employing multiple constraints on the average length of posterior credible intervals of group means

and their difference are derived based on the natural conjugate prior family. Both optimal ratio and fixed ratio designs are explained in this section. We continue in section 4.4 with an illustrative example and two real-life phase II studies for sample-size determination using the proposed Bayesian designs. Finally, a brief discussion is given in section 4.5

4.2 Average Length Criterion

Prior to discussing the Bayesian optimal designs, we first review the ALC criterion developed by Joseph et al. (1995). As discussed in section 4.1, this criterion meets our needs and has practical advantages over the decision-theoretic methods for sample size estimation.

To avoid confusion, some comments on notation are needed. We denote matrices or vectors with boldface letters (i.e., $\boldsymbol{\theta}$), and represent scalars by non-bolded letters (i.e., θ). Similar to many text books, a conditional probability density is denoted by $p(.|.)$ and the notation $p(.)$ is used for a marginal distribution. Moreover, we use n, μ, and $ALC_{\alpha}(.)$ to denote the sample size, mean parameter, and average length of a $100(1-\alpha)\%$ posterior credible interval of a parameter of interest, respectively. Subscript notations C and E are employed to indicate the control and experimental arms, respectively; for example, μ_C stands for the mean parameter of control arm. For the difference between means of two arms, $\Delta = \mu_E - \mu_C$ is used.

Suppose that $\boldsymbol{y} = (y_1, ..., y_n)$ is a vector of n independent and identically distributed data points drawn from a population indexed by $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$. The parameter of interest is a function of $\boldsymbol{\theta}$ denoted by $\mu = \mu(\boldsymbol{\theta})$ with dimension one. If the sampling distribution is $p(\boldsymbol{y}|\boldsymbol{\theta})$, and $p(\boldsymbol{\theta})$ is the chosen prior density, then the posterior distribution of $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})}.$$

In practice, the marginal posterior density for $\mu = \mu(\boldsymbol{\theta})$, denoted by $p(\mu|\boldsymbol{y})$, can be computed from $p(\boldsymbol{\theta}|\boldsymbol{y})$ using simulations.

The ALC criterion fixes α and determines the smallest n by guaranteeing that the expected

length of a $100(1-\alpha)\%$ posterior credible interval for the parameter of interest μ weighted by the marginal $p(\boldsymbol{y})$ does not exceed a pre-specified length l. Mathematically, the minimum n is chosen such that

$$ALC_\alpha(\mu|n) = \int \dots \int l_\alpha(\mu|\boldsymbol{y},n)p(\boldsymbol{y})\,d\boldsymbol{y} \leq l, \tag{4.1}$$

where $l_\alpha(\mu|\boldsymbol{y},n)$ is the length of the $100(1-\alpha)\%$ posterior credible interval of μ given data $\boldsymbol{y}$ with size n. Many times, it is not feasible to evaluate integral (4.1) directly; however, it can be expressed as an expectation of $l_\alpha(\mu|\boldsymbol{y},n)$ over the joint density $p(\mu,\boldsymbol{y})$:

$$\begin{aligned} E_{p(\mu,\boldsymbol{y})}\left[l_\alpha(\mu|\boldsymbol{y},n)\right] &= \int \dots \int l_\alpha(\mu|\boldsymbol{y},n)p(\mu,\boldsymbol{y})\,d\mu\,d\boldsymbol{y} \\ &= \int \dots \int l_\alpha(\mu|\boldsymbol{y},n)\left[\int p(\mu,\boldsymbol{y})\,d\mu\right]d\boldsymbol{y} \\ &= ALC_\alpha(\mu|n). \end{aligned}$$

Therefore, if we draw a large sample $\{(\boldsymbol{y}_s,\mu_s), s=1,\dots,S\}$ from the joint distribution $p(\mu,\boldsymbol{y})$, the integral (4.1) can be estimated using Monte Carlo integration as

$$\widehat{ALC}_\alpha(\mu|n) = \frac{1}{S}\sum_{s=1}^{S} l_\alpha(\mu_s|\boldsymbol{y}_s,n). \tag{4.2}$$

Moreover, instead of employing the highest posterior density region, we derive the length of the credible interval using the central posterior interval due to its easy computation and direct interpretation as the posterior $\alpha/2$ to $1-\alpha/2$ inter-quantile range.

4.3 Bayesian Optimal Designs

In this section, the Bayesian optimal designs that minimize the total sample sizes under multiple constraints based on the ALC criterion is proposed. We start this section by refreshing the association between the exponential family and the natural conjugate prior distributions.

4.3.1 Exponential family and natural conjugate prior

A conjugate model is usually a good starting point for Bayesian analysis since it is easy to understand and simplifies the computation in the sense that it often leads to a closed form posterior distribution. In general, only distributions that belong to the exponential family have natural conjugate prior family (Gelman et al., 2014). The review of the natural conjugate distributions and the exponential family closely follow Gelman et al. (2014).

A family of distributions is called an exponential family if all its probability density functions (pdf) or probability mass functions (pmf) can be expressed as

$$p(y|\boldsymbol{\theta}) = h(y)c(\boldsymbol{\theta})exp\left\{\sum_{j=1}^{k}\eta_j(\boldsymbol{\theta})t_j(y)\right\}, \tag{4.3}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_\omega), \omega \leq k$ is a ω-dimensional vector of parameters. Suppose that $\boldsymbol{y} = (y_1, ..., y_n)$ is a vector of n samples such that the y_i's are independent and identically distributed from an exponential family with pdf or pmf of the form (4.3), the likelihood function of this sample is

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \left\{\prod_{i=1}^{n}h(y_i)\right\}\{c(\boldsymbol{\theta})\}^n exp\left\{\sum_{j=1}^{k}\left[\eta_j(\boldsymbol{\theta})\sum_{i=1}^{n}t_j(y_i)\right]\right\}, \tag{4.4}$$

where the statistic $T(\boldsymbol{y}) = (\sum_{i=1}^{n}t_1(y_i), ..., \sum_{i=1}^{n}t_k(y_i))$ is a sufficient statistic for $\boldsymbol{\theta}$, and the vector $\boldsymbol{\eta} = (\eta_1(\boldsymbol{\theta}), ..., \eta_k(\boldsymbol{\theta}))$ is called the natural parameter. It has been shown that, for any sampling distribution from the exponential family, there always exists a natural conjugate prior family, in which all densities have the same functional form as the likelihood function (Gelman et al., 2014). Therefore, likelihood function (4.4) reveals that the family of natural conjugate prior densities for the exponential family follows the form

$$p(\boldsymbol{\theta}) \propto \{c(\boldsymbol{\theta})\}^{\tau_0} exp\left\{\sum_{j=1}^{k}\eta_j(\boldsymbol{\theta})\tau_j\right\}, \tag{4.5}$$

indexed by known hyperparameters $(\tau_0, ..., \tau_k)$. With this conjugate prior family, the posterior

density for $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \{c(\boldsymbol{\theta})\}^{\tau_0+n} exp\left\{\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}) \left[\tau_j + \sum_{i=1}^{n} t_j(y_i)\right]\right\}. \tag{4.6}$$

Comparing the natural conjugate prior density (4.5) with the likelihood function (4.4), it is noted that hyperparameters can be viewed as prior observations in the context of a natural conjugate model.

4.3.2 Design constraints

Unless otherwise mentioned, $\boldsymbol{y}_C = (y_{1,C}, ..., y_{n_C,C})$ and $\boldsymbol{y}_E = (y_{1,E}, ..., y_{n_E,E})$ denote n_C and n_E independent and identically distributed draws for the control and experimental arms, respectively, where $y_{i,C}$ and $y_{i,E}$ are assumed following two independent exponential family distributions conditional on parameters $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_E$. We additionally assume the joint prior distribution $p(\boldsymbol{\theta}_C, \boldsymbol{\theta}_E)$ is equal to $p(\boldsymbol{\theta}_C)p(\boldsymbol{\theta}_E)$; that is, $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_E$ are independent of each other. Equations (4.4), (4.5), and (4.6) demonstrated in section 4.3.1 suggest that the corresponding posterior distribution for $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_E$ using natural conjugate prior distributions can be expressed as

$$p(\boldsymbol{\theta}_C|\boldsymbol{y}_C) \propto \{c(\boldsymbol{\theta}_C)\}^{\tau_{0,C}+n_C} exp\left\{\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}_C) \left[\tau_{j,C} + \sum_{i=1}^{n} t_j(y_{i,C})\right]\right\},$$

and

$$p(\boldsymbol{\theta}_E|\boldsymbol{y}_E) \propto \{c(\boldsymbol{\theta}_E)\}^{\tau_{0,E}+n_E} exp\left\{\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}_E) \left[\tau_{j,E} + \sum_{i=1}^{n} t_j(y_{i,E})\right]\right\},$$

respectively, where $\left(\tau_{0,C}, ..., \tau_{k,C}\right)$ and $\left(\tau_{0,E}, ..., \tau_{k,E}\right)$ are their corresponding known hyperparameters. We are interested in minimizing the total sample size in terms of the expected length of posterior credible intervals of means and their difference; therefore, the parameters of primary interest are $\mu_C = \mu_C(\boldsymbol{\theta}_C)$, $\mu_E = \mu_E(\boldsymbol{\theta}_E)$, and $\Delta = \mu_E - \mu_C = \mu_E(\boldsymbol{\theta}_E) - \mu_C(\boldsymbol{\theta}_C)$, for which posterior distributions can be derived using simulations by drawing two samples first from $p(\boldsymbol{\theta}_C|\boldsymbol{y}_C)$ and $p(\boldsymbol{\theta}_E|\boldsymbol{y}_E)$ and then computing the values of parameters of interest for each sample point. Using equation (4.1) derived in section 4.2, the total sample size, defined as $N = n_C + n_E$, is minimized

under the following three constraints:

$$ALC_\alpha(\mu_C|n_C) = \int ... \int l_\alpha(\mu_C|\boldsymbol{y}_C, n_C) p(\boldsymbol{y}_C)\, d\boldsymbol{y}_C \leq l_C, \tag{4.7}$$

$$ALC_\alpha(\mu_E|n_E) = \int ... \int l_\alpha(\mu_E|\boldsymbol{y}_E, n_E) p(\boldsymbol{y}_E)\, d\boldsymbol{y}_E \leq l_E, \tag{4.8}$$

$$ALC_\alpha(\Delta|n_C, n_E) = \int ... \int l_\alpha(\Delta|\boldsymbol{y}_C, \boldsymbol{y}_E, n_C, n_E) p(\boldsymbol{y}_C, \boldsymbol{y}_E)\, d\boldsymbol{y}_C\, d\boldsymbol{y}_E \leq l_\Delta. \tag{4.9}$$

It is noted that quantities l_C, l_E, and l_Δ all should be positive since the average length of a posterior interval should be larger than zero. Performing direct calculations on the above three integrals seems intractable; therefore, with known hyperparameters $\left(\tau_{0,C}, ..., \tau_{k,C}\right)$ and $\left(\tau_{0,E}, ..., \tau_{k,E}\right)$ for two arms, fixed sample sizes n_C and n_E, and a specified α level, the following Monte Carlo algorithm is developed to identify n_C and n_E that satisfy constraints (4.7) - (4.9).

1. For $s = 1, ..., S$,

   a. Draw $\boldsymbol{\theta}_C^s$ and $\boldsymbol{\theta}_E^s$ from $p(\boldsymbol{\theta}_C)$ and $p(\boldsymbol{\theta}_E)$, respectively; then draw $\boldsymbol{y}_C^s$ and $\boldsymbol{y}_E^s$ from $p(\boldsymbol{y}_C|\boldsymbol{\theta}_C^s)$ and $p(\boldsymbol{y}_E|\boldsymbol{\theta}_E^s)$, respectively.

   b. Generate M draws of $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_E$ from $p(\boldsymbol{\theta}_C|\boldsymbol{y}_C^s)$ and $p(\boldsymbol{\theta}_E|\boldsymbol{y}_E^s)$, respectively; then compute $\mu_C(\boldsymbol{\theta}_C)$, $\mu_E(\boldsymbol{\theta}_E)$, and $\mu_C(\boldsymbol{\theta}_C) - \mu_E(\boldsymbol{\theta}_E)$ for each draw.

   c. Obtain length of credible intervals $l_\alpha^s(\mu_C|\boldsymbol{y}_C^s, n_C)$, $l_\alpha^s(\mu_E|\boldsymbol{y}_E^s, n_E)$, and $l_\alpha^s(\Delta|\boldsymbol{y}_C^s, \boldsymbol{y}_E^s, n_C, n_E)$ by taking the $\alpha/2$ and $1-\alpha/2$ sample quantiles from M values of μ_C, μ_E, and $\mu_E - \mu_C$ derived in step 2.

2. Left-hand sides of constraints (4.7) - (4.9) can then be approximated using formula (4.2) in

section 4.2 as

$$\widehat{ALC}_\alpha(\mu_C|n_C) = \frac{1}{S}\sum_{s=1}^{S} l_\alpha^s(\mu_C|\boldsymbol{y}_C^s, n_C),$$
$$\widehat{ALC}_\alpha(\mu_E|n_E) = \frac{1}{S}\sum_{s=1}^{S} l_\alpha^s(\mu_E|\boldsymbol{y}_E^s, n_E),$$
$$\widehat{ALC}_\alpha(\Delta|n_C, n_E) = \frac{1}{S}\sum_{s=1}^{S} l_\alpha^s(\Delta|\boldsymbol{y}_C^s, \boldsymbol{y}_E^s, n_C, n_E).$$

It is noted that, under the above Monte Carlo procedure, the lengths of posterior credible intervals are computed based on M posterior samples of two group means and their difference, and each credible interval is evaluated S times. Therefore, fairly good accuracy should be retained as long as the chosen posterior and Monte Carlo sample sizes are large enough. We carried out a small simulation study in order to estimate what values of M and S are sufficiently large for assuring accurate sample size estimates, assuming the prior and likelihood specifications are appropriate. As shown in Figure 4.1, the center of estimated one-to-one ratio total sample size with an exponential outcome is fairly stable as the number of independent runs (i.e., $S = M$) increases. The trial-to-trial variability is relatively large for small $S = M$. It is fairly stable and reaches a reasonably small magnitude when $S = M \geq 3000$. This suggests that the Monte Carlo approach may work well for the ALC criterion with relatively small S and M. In addition, one could further reduce Monte Carlo errors by averaging over a fixed number of parallel simulation trains.

In two-arm phase II clinical trials, subjects can be randomly assigned to two arms at a fixed or flexible allocation ratio. Thus, the proposed Bayesian optimal designs consist of two schemes: the fixed ratio design and the optimal ratio design. Both are discussed in the following two sections.

4.3.3 Fixed ratio design

Sometimes investigators desire to assign patients to the control and experimental arms with a fixed ratio $r = c/e$, where c and e are both positive integers. For example, economic efficiency can potentially be improved by assigning more patients to the less expensive arm. When this is the

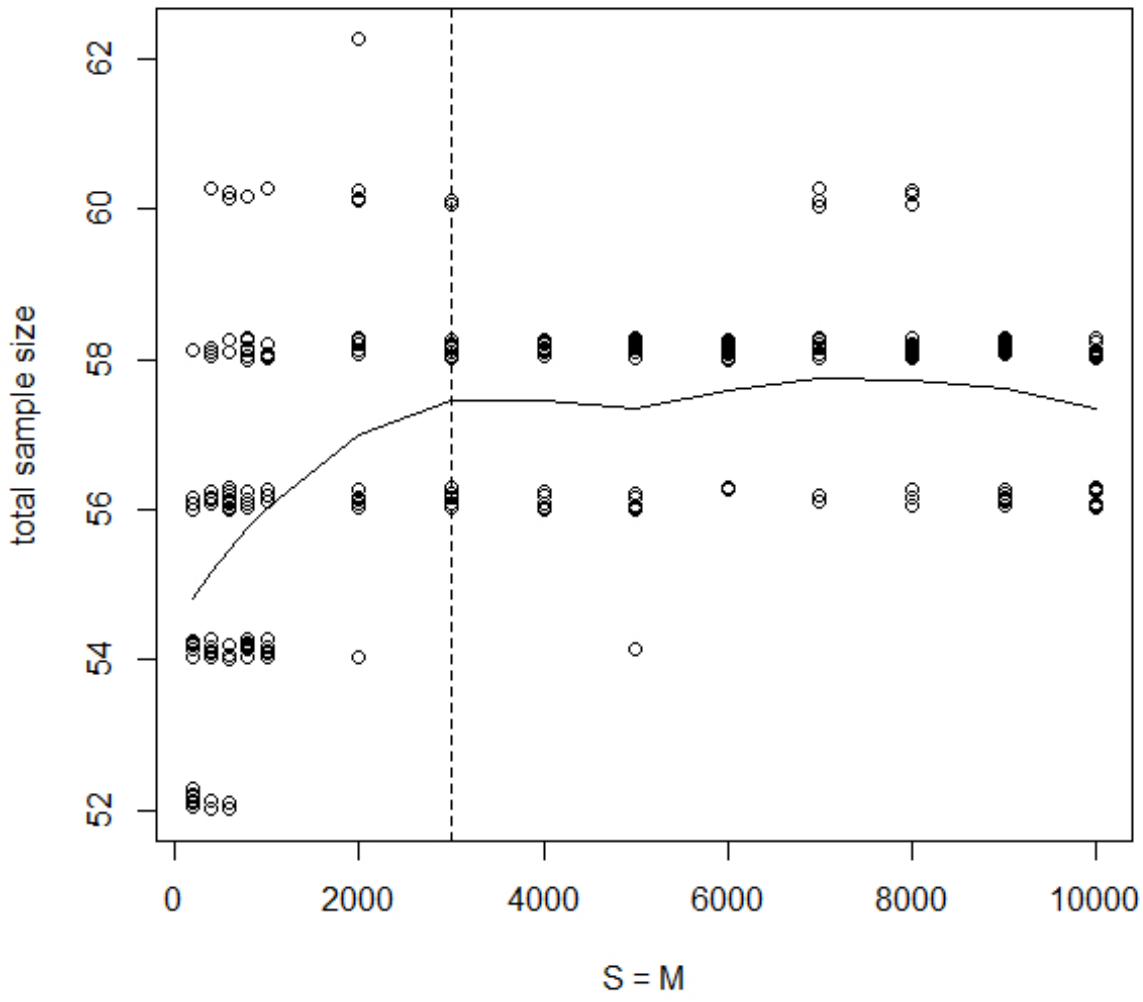

Figure 4.1: The total sample sizes estimated for 1 to 1 design using Monte Carlo approach against $S = M$ for exponential distribution with Inverse-Gamma$(3,40)$ prior for control and Inverse-Gamma(3, 56) prior for experimental arms. The horizontal line corresponds to loess fit and vertical line corresponds to $M = S = 3000$. The point positions have been jittered vertically to reduce overlap.

case, the desired point (n_C, n_E) needs to satisfy the relationship $n_C = rn_E$ as well as constraints (4.7) - (4.9). Under the assumption that the greatest common divisor for c and e is one (i.e., relative prime), Mayo et al. (2010) demonstrated that the set $\{(cd, ed) : d = 1,2,3,...,\infty\}$ contains all two-dimensional points such that their coordinates satisfy the ratio $r = c/e$.

We perform a grid search, a commonly used numerical method for searching for an optimal point, to find the minimum sample size that satisfies design constraints and the fixed ratio $r = c/e$.

To begin with, the crude starting value of d is estimated by applying the Monte Carlo simulation algorithm discussed in section 4.3.2 with a small value of $S = M$ (say, 500). Starting from grid value $d = 1$, we first substitute (cd, ed) for (n_C, n_E) in three integrals under constraints (4.7) - (4.9), and then apply the Monte Carlo algorithm to evaluate each constraint. If any of them fail to meet the corresponding constraint (or constraints), we increase d by 1. This procedure is not terminated until all estimated expected lengths of posterior credible intervals resulted from (cd, ed) are smaller than or equal to pre-specified upper limits. The resulting new d (i.e., d_0) is the estimated starting value. Now we increase the number of independent runs to a larger value ($S = M \geq 3000$) that reduces trial-to-trial sample size variability to a reasonable magnitude. The fixed ratio minimum sample size is derived by applying the following algorithm:

Step 1: Let $d = d_0$ be the starting grid value.

Step 2: With defined c and e, replace (n_C, n_E) in the left-hand side of constraints (4.7) - (4.9) by (cd, ed). Then, apply the Monte Carlo algorithm to approximate integrals under three design constraints. If any lead to values larger than the corresponding pre-specified cutoff levels (i.e., l_C, l_E, and l_Δ), then $d = d + 1$; Otherwise, estimate the three integrals under the design constraints using $d = d - 1$ to evaluate whether current d value is actually larger than the desired minimum.

Step 3: Repeat step 3 until the minimum d is attained. The minimum sample size for $c : e$ fixed ratio design is $(n_C, n_E) = (cd, ed)$.

4.3.4 Optimal ratio design

When there is no limitation on the allocation ratio, the minimum sample size needs to satisfy constraint (4.7) - (4.9) only. Again, the grid search method is conducted to obtain the minimum sample size under the design constraints. However, for the optimal ratio design, the grid point (n_C, n_E) is two-dimensional and the standard grid search method evaluates all combinations of grid values of n_C and n_E. Thus, as the number of values taken by n_C and n_E increases, the number of

grid points inflates considerably and searching speed could be extremely slow. For example, when both n_C and n_E range from 1 to 10 by increments of 1, the total number of grids is $10^2 = 100$. If both n_C and n_E take integer values from 1 to 30, the total number of grids overwhelmingly increases to $30^2 = 900$. Therefore, we propose a rapid grid search method that lightens the computational burden by efficiently reducing the number of grid points.

Figure 4.2 displays the rapid grid search procedure. To begin with, the Monte Carlo simulation procedure and the standard grid search method are conducted to obtain the minimal sample size $(n_C^{(4.7)}, n_E^{(4.8)})$ under constraints (4.7) and (4.8). As discussed previously, a small $S = M$ value may be helpful for targeting an efficient starting value when the standard grid search method is performed. We set the resulting point as the starting grid point, that is, $(n_C^0, n_E^0) = (n_C^{(4.7)}, n_E^{(4.8)})$. Be aware that $(n_C^{(4.7)}, n_E^{(4.8)})$ reaches the minimal sample size if it does not violate constraint (4.9). Now, we add one sample at a time to the experimental arm and evaluate constraint (4.9) for each grid point using the Monte Carlo simulation until the approximated expected length of the posterior credible interval for $(\mu_E - \mu_C)$ is smaller than l_Δ. This gives us an initial optimal candidate (n_C^0, n_E^1). We next add one sample to n_C^0 and withdraw one sample from n_E^1, and again perform the Monte Carlo simulation to estimate the mean length of the posterior credible interval in constraint (4.9). If it is larger than the cutoff level l_Δ, additional samples are added and removed, respectively, to the control and from the experimental arms one at a time until the estimated average length of the posterior credible interval of $(\mu_E - \mu_C)$ satisfies constraint (4.9). We then make subsequent searches by decreasing one sample at a time from the experimental arm until the grid point fails to meet constraint (4.9). This yields the second optimal candidate (n_C^1, n_E^2). Suppose that k optimal candidates $\left\{(n_C^{j-1}, n_E^j), k = 1, ..., k\right\}$ are obtained by repeating the above steps; this procedure is terminated if $n_E^k - 1$ is smaller than n_E^0. We give the algorithm of the proposed rapid grid search method as follows:

Step 1: Use the Monte Carlo approximation and the standard grid search method to derive the minimum sample size under constraints (4.7) and (4.8). Consider this point as the starting grid point and denote it by (n_C^0, n_E^0). If $\widehat{ALC}_\alpha(\Delta | n_C^0, n_E^0) \leq l_\Delta$, then (n_C^0, n_E^0) is the optimal

point under the optimal ratio design; otherwise, go to step 2.

Step 2: For $i = 1, 2, ..., \infty$, evaluate $\widehat{ALC}_\alpha(\Delta|n_C, n_E + i)$ until it is no greater than l_Δ. Suppose $i = I$, then report $(n_C, n_E) = (n_C, n_E + I)$ as an optimal candidate and proceed to step 3.

Step 3: Add samples to the control arm and remove samples from the experimental arm one at a time until the updated $\widehat{ALC}_\alpha(\Delta|n_C, n_E)$ is less than or equal to l_Δ. Then make subsequent searches downwards by withdrawing one sample at a time from the experimental arm until the updated $\widehat{ALC}_\alpha(\Delta|n_C, n_E)$ is greater than l_Δ. The last grid point that gives an estimated average length of no greater than l_Δ is a candidate of optimal point.

Step 4: Repeat step 3 until n_E is less than n_E^0.

For the optimal ratio design, the grid search method may lead to multiple optimal points with the same minimum total sample size but different allocation schemes. During clinical practice, investigators can choose the point that most satisfies their needs. For example, a design that assigns more patients to the less expensive arm may be preferred for the sake of reducing the trial budget.

4.4 Application

In section 4.3, we demonstrated the Bayesian optimal designs for two-arm randomized phase II clinical trials, which optimize the total sample sizes under multiple constraints on the average length of posterior credible intervals of two group means and their difference. In this section, we first briefly discuss strategies for prior elicitation in section 4.4.1. We then continue in section 4.4.2 with an illustrative example that explains implementations of the proposed models on the exponential distribution. Finally, we conclude this section with two real-life phase II clinical trials for which the primary outcome for one trial is normally distributed and the other follows the Bernoulli distribution. R functions are developed and used for computing sample sizes in all examples.

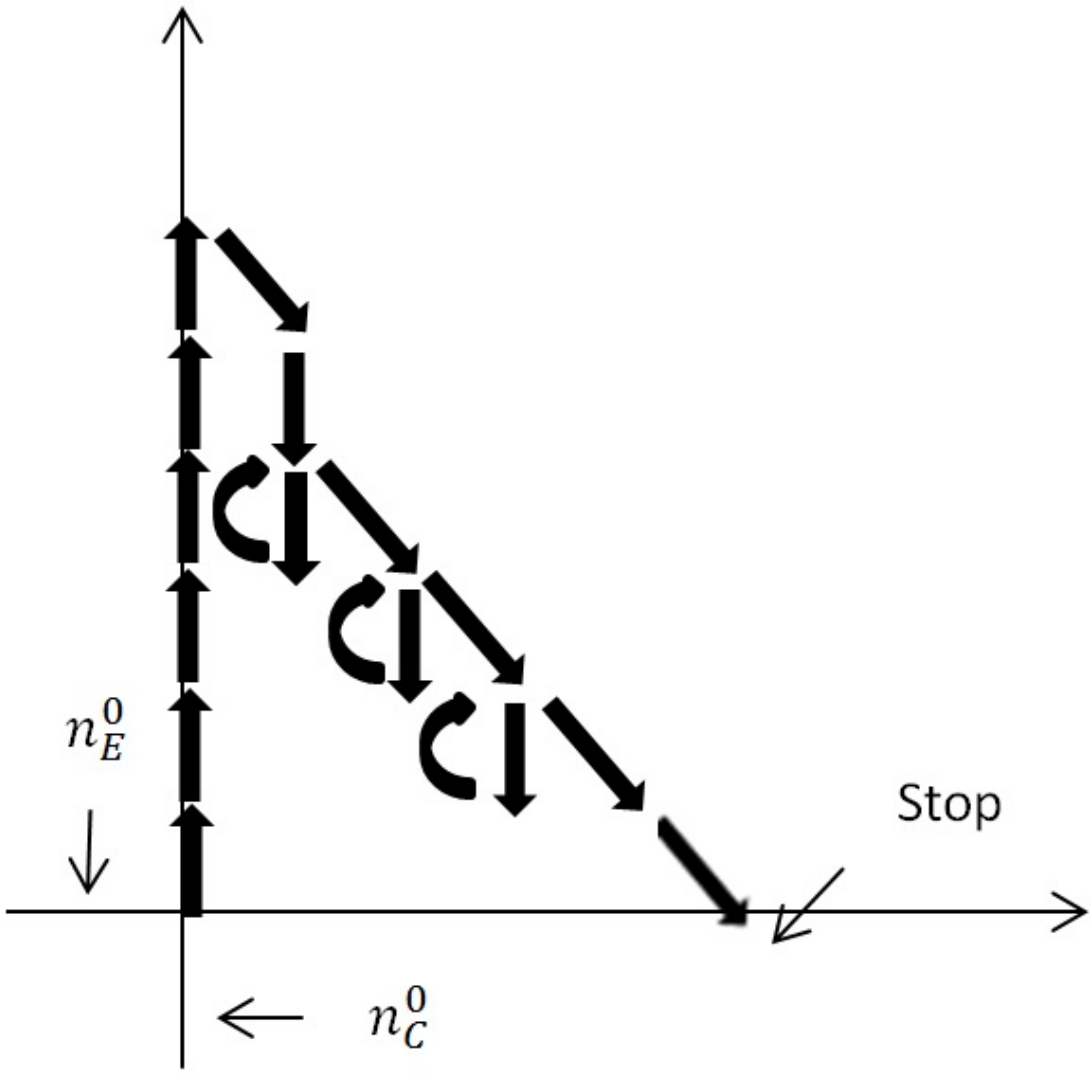

Figure 4.2: Rapid grid search procedure. The searching process stops when the grid value for n_E is less than n_E^0.

4.4.1 Prior elicitation

In this subsection, we briefly illustrate our general approach for eliciting natural conjugate prior distributions. Depending on the amount of information provided from the prior distribution, priors can be classified into informative, weakly informative, and non-informative relative to the likelihood. More specifically, informative priors express relatively small prior uncertainty about parameters; weakly informative priors convey relatively less certain information about likely values of

parameters; and the non-informative priors reflect high prior uncertainty (Gelman et al., 2014).

Many authors have discussed approaches of prior elicitation, including Spiegelhalter et al. (2004), O'Hagan et al. (2006), and Johnson et al. (2010). As pointed out by Spiegelhalter et al. (2004), these different approaches can be classified into four categories: (1) informal discussion, (2) structural interviewing and formal pooling of opinion, (3) structured questionnaires, and (4) computer-based elicitation. In this paper, we employ the method discussed in Mayo & Gajewski (2004) which includes two steps for eliciting priors. First, the center value of the chosen prior is elicited by asking a clinician questions such as, " What is the most likely average value that you expect to observe for the primary endpoint?" The response to this question can correspond to the prior mean. Other commonly used center measurements include the median and mode. The second step is to elicit the uncertainty in the center value by asking clinicians questions such as, "How uncertain is the center value you provided?" Depending on different answers provided from clinicians, the uncertainly in center value can be elicited in terms of the prior sample size or the width of a $100(1-\alpha)\%$ credible interval from the prior distribution. The computer plot of the pdf or pmf of the corresponding conjugate distribution may be helpful for clinicians to conceptualize the prior (Thall & Simon, 1994).

Gelman et al. (2014) demonstrated that hyperparameters of natural conjugate priors can be interpreted as additional observations. Thus, we elicit prior densities in terms of the prior mean and prior sample size. For example, informative priors are built by fixing the prior mean and choosing a large prior sample size relative to the size of the experiment. Conversely, weakly informative priors are constructed using a small prior sample size relative to the size of the experiment. Following this idea, the hyperparameters for both control and experimental arms are obtained by specifying the prior means and the prior sample sizes which parameterize two prior distributions to properly reflect the experts' opinion. The pre-specified prior information and the solutions to these simultaneous equations fully specify the prior distributions.

4.4.2 Exponential endpoint

The exponential distribution is a natural and commonly used model in phase II clinical trials when the data are in the form of waiting time. For instance, the primary clinical endpoint in many phase II oncology trials is a time-to-event outcome such as the disease-free or overall survival.

4.4.2.1 InverseGamma-Exponential model

When the observations $\boldsymbol{y}_g = (y_{1,g}, ..., y_{n_g,g})$ provided for arm g under the experiment of consideration are independent and identically exponentially distributed with scale parameter θ_g, $g = C, E$, the likelihood can be written in the exponential family form (4.4) as

$$p(\boldsymbol{y}_g|\theta_g) = \left(\frac{1}{\theta_g}\right)^{n_g} exp\left\{\frac{-\sum_{i=1}^{n_g} y_{i,g}}{\theta_g}\right\},$$

where $c(\theta_g) = 1/\theta_g$, $\eta_1(\theta_g) = -1/\theta_g$ and $\sum_{i=1}^{n_g} t_1(y_{i,g}) = \sum_{i=1}^{n_g} y_{i,g}$ with $k = \omega = 1$. Equation (4.5) suggests that the natural conjugate prior distribution for θ_g is

$$p(\theta_g) \propto \left(\frac{1}{\theta_g}\right)^{\alpha_g+1} exp\left\{\frac{-\beta_g}{\theta_g}\right\},$$

where $\tau_{0,g} = \alpha_g + 1$ and $\tau_{1,g} = \beta_g$. This is an Inverse-Gamma density with shape parameter α_g and rate paramter β_g. Using equation (4.6), the corresponding posterior distribution for θ_g is

$$\theta_g|\boldsymbol{y}_g \sim \text{Inverse-Gamma}(\alpha_g + n_g, \beta_g + \sum_{i=1}^{n_g} y_{i,g}).$$

According to Gelman et al. (2014), the above two hyperparameters, in some sense, can be explained as $\alpha_g + 1$ exponential observations with total waiting time β_g. Recall that the mean parameter of an $\exp(\theta_g)$ outcome is $\mu_g(\theta_g) = \theta_g$, $g = C, E$; therefore, the three design constraints under which the total sample size is minimized are: $ALC_\alpha(\theta_C|n_C) \leq l_C$, $ALC_\alpha(\theta_E|n_E) \leq l_E$, and $ALC_\alpha(\Delta|n_C, n_E) = ALC_\alpha(\theta_E - \theta_C|n_C, n_E) \leq l_\Delta$.

4.4.2.2 Illustrative example

We first elicit all four hyperparameters, $\alpha_C, \beta_C, \alpha_E$, and β_E, in order to perform the Monte Carlo simulation. For illustrative purposes, the InverseGamma-Exponential model is carried out with $E(\theta_C) = \beta_C/(\alpha_C - 1)$, the prior mean of the control arm, equal to 20 and 30, and $E(\theta_E) = \beta_E/(\alpha_E - 1)$, the prior mean of the experimental arm, equal to 24 and 28 when $E(\theta_C) = 20$, and equal to 34 and 38 when $E(\theta_C) = 30$. In order to study the influence of prior specification on sample size estimates, we place both weakly informative and informative priors on θ_C and θ_E. More specifically, the prior sample sizes of both arms are 4 for the weakly informative model, and are 12 for the informative model. Suppose that the estimated total sample size of an experiment is N, the ratios $8/(8+N)$ and $24/(24+N)$, in some sense, can be viewed as the percentages of information contributed by the weakly informative and informative prior distributions, respectively. Finally, regarding the average length cutoff levels for $\alpha = 0.05$, we set $l_C = 21.5$, $l_E = 21.5$, and $l_\Delta = 27$.

The computed minimum sample sizes under the one-to-one, one-to-two, and optimal ratios are reported in Table 4.1. Suppose that one has an actual phase II study where the standard of care has a prior mean of $E(\theta_C) = 20$, and you expect a new treatment to have a prior mean of $E(\theta_E) = 28$. For the informative model, the solution to the prior Inverse-Gamma parameters are $\alpha_C = 11$, $\beta_C = 200$, $\alpha_E = 11$, and $\beta_E = 280$. Under the constraints used in Table 4.1, the optimal ratio design results in a total of 34 subjects; the one-to-one design results in 18 subjects in each arm (36 total subjects); and the one-to-two design results in 12 subjects in the standard of care arm and 24 subjects in the experimental arm (36 total subjects). These imply that approximately 40% information is contributed by the informative prior distributions under each design consideration. Notice that the optimal ratio design produces several different allocation schemes. If investigators want to assign more patients to the more promising experimental arm, the allocation scheme $(11, 23)$ may be chosen. In Table 4.1, we also calculate the optimal sample sizes under the weakly informative model. Not surprisingly, the resulting sample sizes are larger than the ones under the informative model. Consider again the sample size calculation in the case of $E(\theta_C) = 20$ and

Table 4.1: Bayesian optimal sample sizes for $l_C = l_E = 21.5$ and $l_\Delta = 27$ under 1 to 1, 1 to 2, and optimal randomization allocation ratios for exponential distribution.

$E(\mu_C)$	$E(\mu_E)$	Prior	1 to 1	1 to 2	Optimal	(n_C, n_E)
20	24	01	26	27	26	(9,17) (10,16) (11,15) (12,14) (13,13) (14,12)
20	28	01	36	36	34	(11,23) (12,22) (13,21) (14,20) (15, 19)
30	34	01	74	78	74	(30,44) (31,43) (32,42) (33,41) (34,40) (35,39) (36,38) (37,37) (38,36)
30	38	01	86	90	85	(35,50) (36,49) (37,48) (38,47) (39,46)
20	24	02	46	48	45	(18,27) (19,26) (20,25) (21,24) (22,23) (23,22)
20	28	02	54	57	53	(21,32) (22,31) (23,30) (44,29)
30	34	02	96	105	96	(42,54) (43,53) (44,52) (45,51) (46,50) (47,49) (48,48)
30	38	02	110	114	109	(44,65) (45,64) (46,63) (47,62) (48,61) (49,60) (50,59) (51,58) (52,57) (53,56)

Column "Prior" corresponds to different prior distributions where "01" represents the informative prior (prior sample sizes of both arms are 12) and "02" represents the weakly informative prior (prior sample sizes of both arms are 4).

$E(\theta_E) = 28$. When the prior sample sizes of both arms are 4, the solution to the hyperparameters are $\alpha_C = 3$, $\beta_C = 40$, $\alpha_E = 3$, and $\beta_E = 56$. Using the Monte Carlo algorithm, the optimal ratio design, one-to-one, and one-to-two designs would, respectively, require enrolling 53, 54, and 57 subjects, implying that approximately 10% information is contributed by the weakly informative prior distributions under each design consideration.

In addition, we contrast the frequentist optimal designs from Jiang et al. (2014) with the proposed Bayesian optimal designs when the outcome is exponentially distributed. In the context of Bayesian optimal designs, we seek the minimal sample size to ensure that the 95% posterior credible intervals for means of control and experimental arms and their difference have lengths no wider than l_C, l_E, and l_Δ, respectively. Alternatively to 95% posterior credible intervals, the optimal

Table 4.2: Optimal sample sizes for $l_C = l_E = 21.5$ and $l_\Delta = 27$ under 1 to 1, 1 to 2, and optimal randomization allocation ratios for exponential distribution using frequentist approaches (Jiang et al., 2014).

μ_C	μ_E	1 to 1	1 to 2	Optimal	(n_C, n_E)
20	24	42	45	41	(20,21) (19,22) (18,23)
20	28	54	51	49	(22,27) (21,28) (20,29) (19,30)
30	34	88	96	87	(44,43) (43,44) (42,45) (41,46) (40,47) (39,48) (38,49) (37,50)
30	38	100	105	98	(47,51) (46,52) (45,53) (44,54) (43,55) (42,56) (41,57) (40,58)

frequentist approaches search for the total sample size that leads to 95% confidence intervals for the control and experimental means and their difference with lengths being narrower than or equal to l_C, l_E, and l_Δ, respectively. Unlike the Bayesian methods, the frequentist models do not take the uncertainty in values of θ_C and θ_E into account. Table 4.2 presents the sample sizes based on the frequentist designs in which, for comparison purposes, specified values of θ_C and θ_E are chosen to match prior means of these two parameters listed in Table 4.1. Similarly, constraint values of $l_C = l_E = 21.5$ and $l_\Delta = 27$, the same as those employed by the Bayesian methods, are used for computing the frequentist sample sizes. Observe that the sample sizes displayed in Table 4.2 are smaller than the ones obtained from the weakly informative model but are larger than the ones based on the informative priors for all one-to-one, one-to-two, and optimal ratios.

4.4.3 Normal endpoint

The normal distribution is often employed in designs where the original continuous outcome is of interest rather than the categorization of this continuous variable. For example, the tumor size ratio, instead of the response rate, is sometimes treated as the primary endpoint. Lavin (1981) demonstrated that the log tumor size ratio approximately followed the normal distribution.

4.4.3.1 Normal-Inverseχ^2 model

Under the two-parameter normal sampling model, we have $\boldsymbol{\theta}_g = (\mu_g, \sigma_g^2)$, $g = C, E$. The likelihood function in the exponential family form (4.4) for a vector $\boldsymbol{y}_g$ of n_g independent and identical normal distributions with unknown mean μ_g and unknown variance σ_g^2 is

$$p(\boldsymbol{y}_g|\boldsymbol{\theta}_g) \propto \left\{ \left(\frac{1}{\sigma_g^2} \right)^{\frac{1}{2}} \right\}^{n_g} exp \left\{ \frac{\mu_g}{\sigma_g^2} \sum_{i=1}^{n_g} y_{i,g} + \frac{-1}{2\sigma_g^2} \sum_{i=1}^{n_g} \left(y_{i,g}^2 + \mu_g^2 \right) \right\},$$

where $c(\boldsymbol{\theta_g}) = (1/\sigma_g^2)^{1/2}$, $\eta_1(\boldsymbol{\theta}_g) = \mu_g/\sigma_g^2$, $\eta_2(\boldsymbol{\theta}_g) = -1/(2\sigma_g^2)$, $\sum_{i=1}^{n_g} t_1(\boldsymbol{y}_{i,g}) = \sum_{i=1}^{n_g} y_{i,g}$, and $\sum_{i=1}^{n_g} t_2(\boldsymbol{y}_{i,g}) = \sum_{i=1}^{n_g} (y_{i,g}^2 + \mu_g^2)$ with $k = \omega = 2$. Recall from equation (4.5) that the natural conjugate prior for $\boldsymbol{\theta}_g$ has the same functional form as the above likelihood; therefore, the family of conjugate prior densities looks like

$$p(\boldsymbol{\theta_g}) \propto \left\{ \left(\frac{1}{\sigma_g^2} \right)^{\frac{1}{2}} \right\}^{\alpha_g + 3} exp \left\{ \frac{\mu_g}{\sigma_g^2} \left[\kappa_g \zeta_g \right] + \frac{-1}{2\sigma_g^2} \left[\alpha_g \beta_g + \kappa_g \mu_g^2 + \kappa_g \zeta_g^2 \right] \right\},$$

where $\tau_{0,g} = \alpha_g + 3$, $\tau_{1,g} = \kappa_g \zeta_g$, and $\tau_{2,g} = \alpha_g \beta_g + \kappa_g \mu_g^2 + \kappa_g \zeta_g^2$. After some algebra, it can be shown that

$$p(\boldsymbol{\theta_g}) \propto \left(\frac{1}{\sigma_g^2} \right)^{\frac{\alpha_g}{2}+1} exp \left\{ -\frac{\alpha_g \beta_g}{2\sigma_g^2} \right\} \left(\frac{1}{\sigma_g^2} \right)^{\frac{1}{2}} exp \left\{ -\frac{\kappa_g}{2\sigma_g^2} (\mu_g - \zeta_g)^2 \right\}.$$

Therefore, the corresponding conjugate prior density for $\boldsymbol{\theta}_g = (\mu_g, \sigma_g^2)$ is a mixture of the marginal prior distribution $\sigma_g^2 \sim$ Inverse-$\chi^2(\alpha_g, \beta_g)$ and the conditional prior distribution $\mu_g|\sigma_g^2 \sim$ N$(\zeta_g, \sigma_g^2/\kappa_g)$. We label this the Normal-Inv$\chi^2(\zeta_g, \sigma_g^2/\kappa_g; \alpha_g, \beta_g)$, where ζ_g and σ_g^2/κ_g represent location and scale of μ_g with κ_g prior measurements, and the last two parameters can be viewed as α_g prior observations with β_g average squared deviation (Gelman et al., 2014). After using equation (4.6) and some algebraic manipulation, it can be shown that the joint posterior density for

$\boldsymbol{\theta}_g = (\mu_g, \sigma_g^2)$ is

$$\boldsymbol{\theta_g}|\boldsymbol{y}_g \sim \text{Normal-Inv}\chi^2\left(\zeta_{n,g}, \frac{\beta_{n,g}}{\kappa_{n,g}}; \alpha_{n,g}, \beta_{n,g}\right),$$

where

$$\begin{aligned}
\zeta_{n,g} &= \frac{\kappa_g}{\kappa_g + n_g}\zeta_g + \frac{n_g}{\kappa_g + n_g}\bar{y}_g, \\
\kappa_{n,g} &= \kappa_g + n_g, \\
\alpha_{n,g} &= \alpha_g + n_g, \\
\alpha_{n,g}\beta_{n,g} &= \alpha_g\beta_g + \sum_{i=1}^{n_g}(y_{i,g} - \bar{y}_g)^2 + \frac{\kappa_g n_g}{\kappa_g + n_g}(\bar{y}_g - \zeta_g)^2.
\end{aligned}$$

Notice that both the prior information and the information provided by the data are contained in the parameters of the joint posterior distribution. For instance, the parameter $\zeta_{n,g}$ can be explained as a weighted average of the prior mean ζ_g and the sample mean $\bar{y}_g$. Furthermore, because the mean parameter of a normally distributed endpoint is $\mu_g(\boldsymbol{\theta}_g) = \mu_g$, $g = C, E$, the design constraints are $ALC_\alpha(\mu_C|n_C) \leq l_C$, $ALC_\alpha(\mu_E|n_E) \leq l_E$, and $ALC_\alpha(\Delta|n_C, n_E) = ALC_\alpha(\mu_E - \mu_C|n_C, n_E) \leq l_\Delta$.

4.4.3.2 Illustrative example: change in tumor size as endpoint

A recently published randomized phase II clinical study of the effect of sorafenib and erlotinib in non-small-cell lung cancer (Karrison et al., 2007) treats tumor size as a continuous variable for assessing antitumor activity. The control arm of this study is 150 mg erlotinib daily plus placebo, and the two treatment arms are 150 mg erlotinib daily plus 200 mg and 400 mg sorafenib twice daily, respectively. The ratio of tumor size at 8 months after treatment to that at baseline is the primary endpoint, and the author assumes that the log of this ratio comes from a normal distribution. In this paper, we use the control arm and the 150 mg erlotinib plus 400 mg sorafenib to illustrate the proposed optimal designs. Based upon four previous single-arm studies of erlotinib or sorafenib, this study assumes that the mean log ratio is 0.05 for the placebo control arm and is -0.13 for the chosen experimental arm. In addition, this trial assumes an equal variance of 0.1197

for the log ratio by pooling the variance estimates from the four previous trials. For more details about this phase II cancer trial, see Karrison et al. (2007).

In order to use the ALC criterion, we set the cutoff levels for 95% credible intervals ($\alpha = 0.05$) to be $l_C = l_E = 0.2$ and $l_\Delta = 0.28$. Following the elicitation strategy discussed in section 4.4.1 and using the information provided in the above phase II trial, we assume that $E(\mu_C|\sigma_C^2) = \zeta_C = 0.05$ and $E(\sigma_C^2) = (\alpha_C\beta_C)/(\alpha_C - 2) = 0.1197$ for the control arm, and $E(\mu_E|\sigma_E^2) = \zeta_E = -0.13$ and $E(\sigma_E^2) = (\alpha_E\beta_E)/(\alpha_E - 2) = 0.1197$ for the experimental arm. Moreover, we let $\alpha_C = \alpha_E = 5$ which, in some sense, can be viewed as information of five prior observations provided for σ_C^2 and σ_E^2. For comparative purposes, we report the optimal total sample sizes using both weekly informative and informative priors for mean parameters μ_C and μ_E. In more detail, we let $\kappa_C = \kappa_E = 0.001$ for the weakly informative model. This means that nearly no prior information is provided for μ_C and μ_E. In this case, the optimal ratio design requires enrolling 85 total subjects where allocation schemes include $(40,45)$, $(41,44)$, $(42,43)$, $(43,42)$, $(44,41)$, and $(45,40)$; the one-to-one design leads to 86 total subjects among which each arm contains 43 subjects; and the one-to-two design results in 120 total subjects where 40 individuals are assigned to the control and 80 individuals are assigned to the experimental arm. On the contrary, an informative model with $\kappa_C = \kappa_E = 10$ results in 66, 66, and 90 total subjects for the optimal, one-to-one, and one-to-two ratios, respectively. Similar to the weakly informative model, the optimal ratio design using informative priors produces several allocation schemes including $(30,36)$, $(31,35)$, $(32,34)$, $(33,33)$, $(34,32)$, $(35,31)$, and $(36,30)$. As expected, the total sample sizes produced under weakly informative model are larger than the ones obtained from informative model. Figure 4.3 shows the one-to-one ratio sample size determination for the normal endpoint based on informative priors. Further, it should be noticed that, in the context of Normal-Inverseχ^2 model, the resulting total sample size is irrelevant to the location of center value of μ_C and μ_E but affected by the uncertainty in the center values. This is because the variance and mean of the normal distribution are independent of each other.

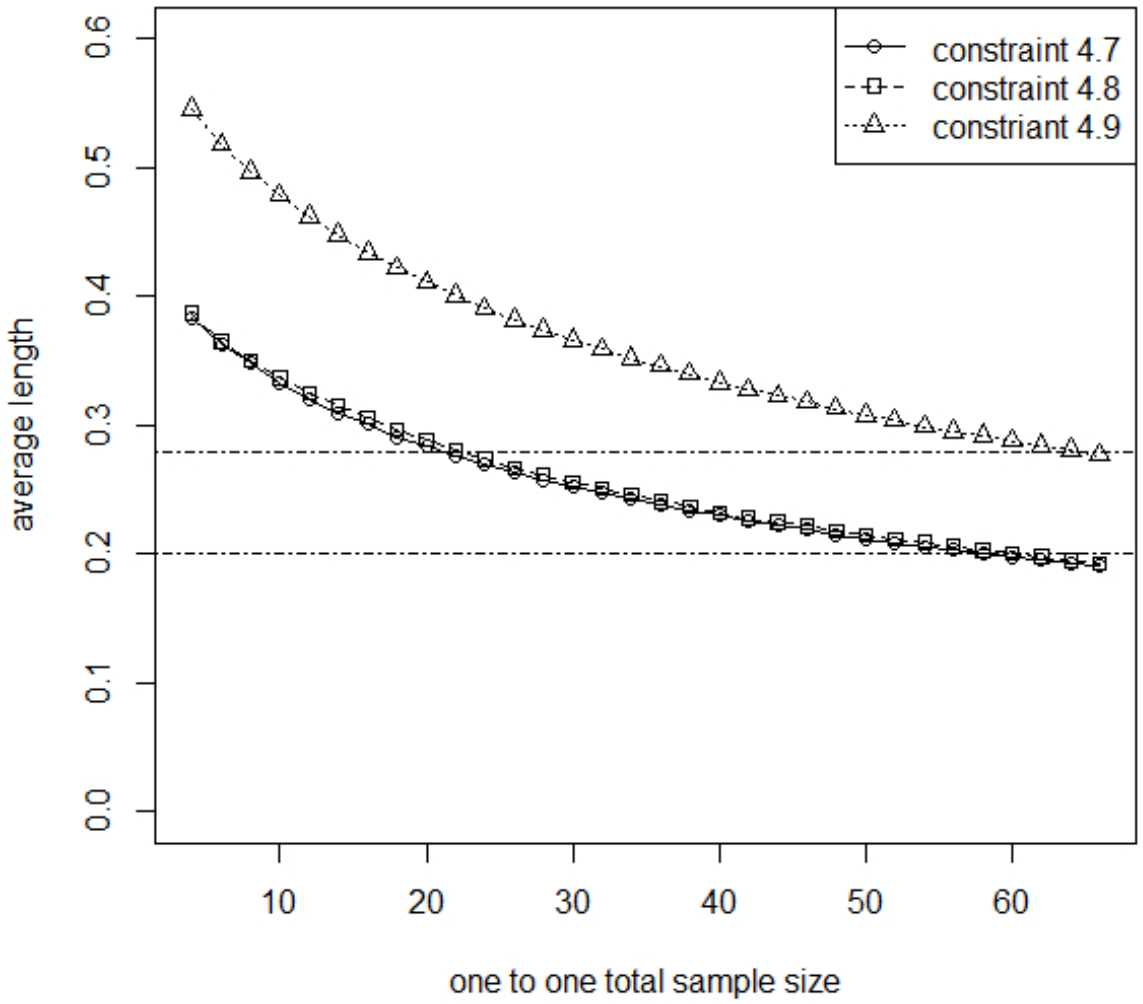

Figure 4.3: Total sample size determination for 1 to 1 design by ALC for normal distribution with $\zeta_C = 0.05$, $\kappa_C = 10$, $\alpha_C = 5$, and $\beta_C = 0.07182$ for control arm, and $\zeta_E = -0.13$, $\kappa_E = 10$, $\alpha_E = 5$, and $\beta_E = 0.07182$ for experimental arm. The horizontal line above corresponds to $l_\Delta = 0.28$ and the horizontal line below corresponds to $l_C = l_E = 0.2$. In the legend, constraint 4.7 stands for $ALC_\alpha(\mu_C|n_C) \leq l_C$, constraint 4.8 stands for $ALC_\alpha(\mu_E|n_E) \leq l_E$, and constraint 4.9 stands for $ALC_\alpha(\Delta|n_C, n_E) \leq l_\Delta$.

4.4.4 Bernoulli endpoint

The Bernoulli distribution is one of the most commonly used models in phase II oncology clinical trials where the primary outcome, for example, may be whether a partial or complete tumor response is observed from a patient or whether a patient has survived one year later after receiving treatment.

4.4.4.1 Beta-Bernoulli model

We let $\boldsymbol{y}_g = (y_{1,g}, ..., y_{n_g,g})$ be a vector of independent and identically distributed Bernoulli draws for arm g with probability parameter θ_g, $g = C, E$. Then the likelihood function for this family can be written in the exponential family form (4.4) as

$$p(\boldsymbol{y}_g|\theta_g) = (1-\theta_g)^{n_g} exp\left\{log\left(\frac{\theta_g}{1-\theta_g}\right)\sum_{i=1}^{n_g} y_{i,g}\right\},$$

where $c(\theta_g) = 1-\theta_g$, $\eta_1(\theta_g) = log\{\theta_g/(1-\theta_g)\}$, and $\sum_{i=1}^{n_g} t_1(y_{i,g}) = \sum_{i=1}^{n_g} y_{i,g}$ with $k = \omega = 1$. Equation (4.5) reveals that the natural conjugate prior density should be of the same form as the likelihood function; therefore, we parametrize the prior distribution as

$$p(\theta_g) \propto (1-\theta_g)^{\alpha_g+\beta_g-2} exp\left\{log\left(\frac{\theta_g}{1-\theta_g}\right)(\alpha_g - 1)\right\},$$

where $\tau_{0,g} = \alpha_g + \beta_g - 2$ and $\tau_{1,g} = \alpha_g - 1$. After several steps simple algebra, this natural conjugate prior density can be equivalently written as

$$p(\theta_g) \propto \theta_g^{\alpha_g-1}(1-\theta_g)^{\beta_g-1},$$

which is a beta density with parameters α_g and β_g. Using equation (4.6), the corresponding posterior distribution for θ_g is therefore

$$\theta_g \sim \text{Beta}\left(\alpha_g + \sum_{i=1}^{n_g} y_{i,g}, \beta_g + n_g - \sum_{i=1}^{n_g} y_{i,g}\right).$$

Be aware that the above two hyperparameters can be explained as $\alpha_g - 1$ prior successes and $\beta_g - 1$ prior failures for a particular sequence of $\alpha_g + \beta_g - 2$ prior experiments (Gelman et al., 2014). In addition, in the context of the Beta-Bernoulli model, the parameter of interest is $\mu_g(\theta_g) = \theta_g$, $g = C, E$; thus the three constraints under which the total sample size is minimized are $ALC_\alpha(\theta_C|n_C) \leq l_C$, $ALC_\alpha(\theta_E|n_E) \leq l_E$, and $ALC_\alpha(\Delta|n_C, n_E) = ALC_\alpha(\theta_E - \theta_C|n_C, n_E) \leq l_\Delta$.

4.4.4.2 Illustrative example: MK-0646 IMPACT study

A recent phase II clinical trial at the University of Kansas Medical Center, the MK-0646 IMPACT study, aimed to compare the response rates of two chemotherapy treatments in non-small cell lung cancer. The standard of care of this study is pemetrexed and cisplatin, and the experimental arm is the combination of pemetrexed/cisplatin with MK-0646. The number of patients was found using the optimal randomized phase II design based upon multiple constraints (Mayo et al., 2010; Jiang et al., 2014). More specifically, they seek the minimal sample size to guarantee that the length of a 90% confidence interval for the response rate in each arm is no wider than 0.33, as well as a 90% confidence interval for the difference between the response rates has a length no greater than 0.4. Moreover, the expected response rate of 0.3 is used for the control arm, and 0.45 is the expected rate for the experimental arm. As a result, the optimal, one-to-one, and two-to-one ratios lead to 62, 62, and 69 total patients, respectively.

Instead of confidence intervals as used in the original study design, the proposed Bayesian approaches minimize the total sample sizes based on posterior credible intervals. Using the prior information provided above, we assume that $E(\theta_C) = \alpha_C/(\alpha_C+\beta_C) = 0.3$ and $E(\theta_E) = \alpha_E/(\alpha_E+\beta_E) = 0.45$. Now suppose that we have a strong belief in the prior knowledge about θ_C (standard care) but are not confident in the prior information for θ_E (new treatment). This may be reflected by placing an informative prior on the control arm but a weakly informative prior on the experimental arm. More specifically, we let $\alpha_C+\beta_C = 12$ which corresponds to 10 prior observations, and $\alpha_E+\beta_E = 2$ which corresponds to no prior observation. In this case, solutions to the hyperparameters are $\alpha_C = 3.6$, $\beta_C = 8.4$, $\alpha_E = 0.9$, and $\beta_E = 1.1$. In addition, we use the same cutoff levels as those used in the frequentist methods; that is, $l_C = l_E = 0.33$ and $l_\Delta = 0.4$. Under the specified constraints, the optimal ratio design leads to 34 total patients; the one-to-one design requires 36 total patients, 18 patients in each arm; and the two-to-one design results in 42 total patients where 28 patients are randomized to the control and 14 patients are randomized to the experimental arm. As usual, several allocation schemes are produced by the optimal ratio design including $(11,23)$, $(12,22)$, $(13,21)$, $(14,20)$, and $(15,19)$. For comparison purposes, we also compute the sample

sizes by placing non-informative priors on both study groups. As a result, the non-informative model with $\alpha_C = \beta_C = \alpha_E = \beta_E = 1$ leads to 41, 42, and 45 total subjects for optimal, one-to-one, and two-to-one ratios. In this case, the allocation schemes of the optimal ratio design contain $(18,23)$, $(19,22)$, $(20,21)$, $(21,20)$, $(22,19)$, and $(23,18)$. Not surprisingly, we see that the sample sizes under non-informative model are larger than the ones obtained from more informative priors. Unlike previous examples, however, both prior settings lead to much smaller sample sizes than those obtained from the frequentist approaches. This may be because the flat non-informative priors still provide some information by restricting θs to values between 0 and 1. Figure 4.4 shows the sample size determination of one-to-one ratio design for binary outcomes based on the informative priors.

4.5 Discussion

In this paper, we develop Bayesian optimal designs and discuss their use for designing two-arm randomized phase II clinical trials with endpoints from distributions in the exponential family. The proposed designs minimize the sample sizes by placing multiple constraints on the average length of posterior credible intervals of two group means and their difference. We employ the ALC criterion to determine our sample size due to its practical advantage that the sample size can be calculated without specifying a loss or utility function. Since it is not feasible to present a general closed form formula for ALC, we estimate the average length of posterior credible intervals and the sample size using the Monte Carlo method. Our simulation study shows that the Monte Carlo method performs well with relatively small values of S and M. Moreover, we develop a rapid grid search method to reduce the computational burden.

Unlike the optimal designs proposed by Mayo et al. (2010) and Jiang et al. (2014) using frequentist approaches, the proposed Bayesian methods can take uncertainty in parameters of interest into account through prior distributions. We focus on the natural conjugate prior family because many probability distributions from the exponential family, such as the Bernoulli, normal, Poisson, and exponential, are important in the practice of phase II clinical trials and, according to Gelman

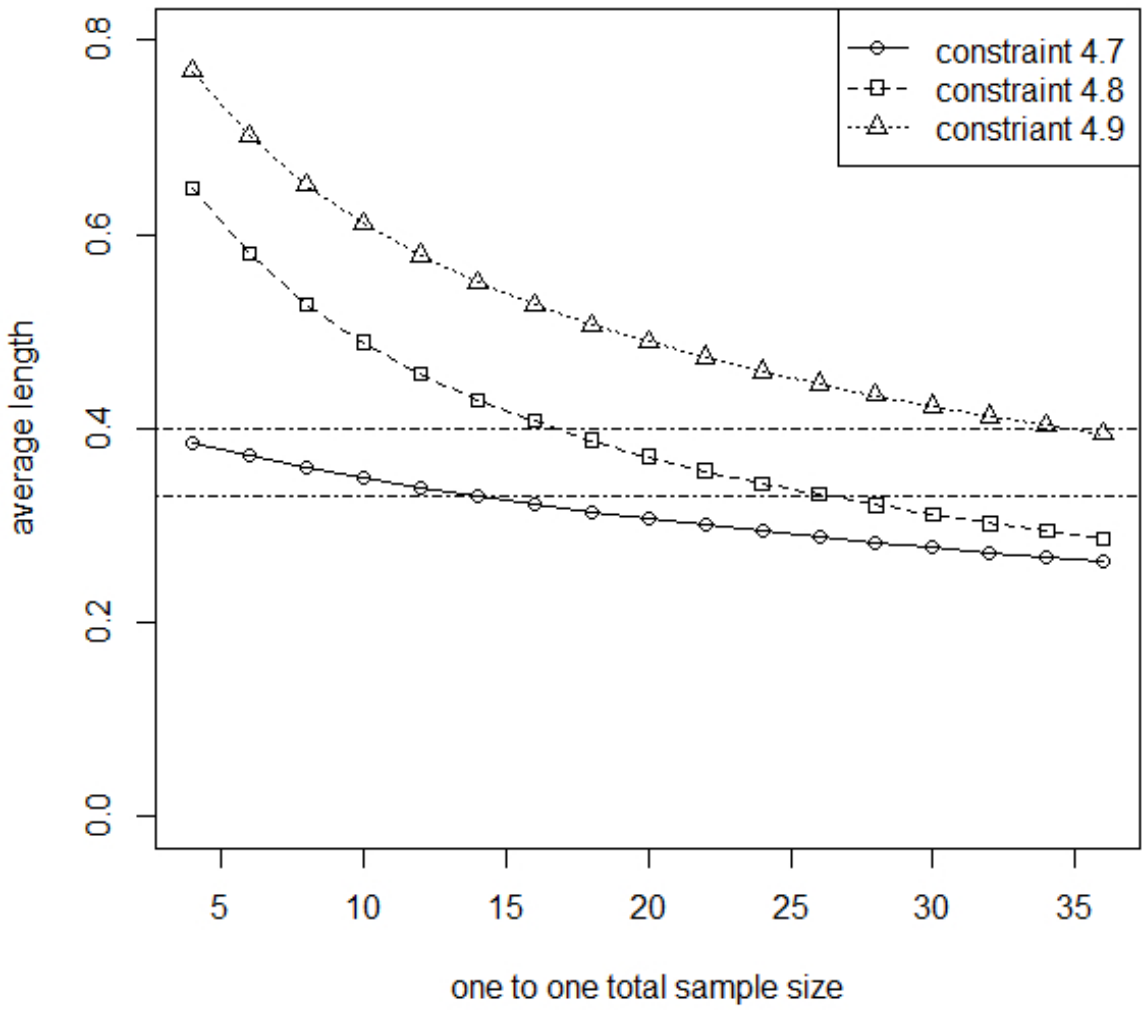

Figure 4.4: Total sample size determination for 1 to 1 design by ALC for Bernoulli distribution with Beta(3.6, 8.4) prior for control arm and Beta(0.9, 1.1) prior for experimental arm. The horizontal line above corresponds to $l_\Delta = 0.4$ and the horizontal line below corresponds to $l_C = l_E = 0.33$. In the legend, constraint 4.7 stands for $ALC_\alpha(\mu_C|n_C) \leq l_C$, constraint 4.8 stands for $ALC_\alpha(\mu_E|n_E) \leq l_E$, and constraint 4.9 stands for $ALC_\alpha(\Delta|n_C, n_E) \leq l_\Delta$.

et al. (2014), the only class of distributions that has natural conjugate prior distributions is the exponential family. Several other advantages associated with the conjugate family include: the posterior distribution of conjugate models has a closed form density so we can easily derive quantities of interest analytically; and the hyperparameters of conjugate prior distributions are easy to understand because, in some sense, they can be interpreted as prior observations (Gelman et al., 2014). However, one limitation is that some prior beliefs cannot be fully described by conjugate

priors. For example, a beta prior cannot be bimodal. Moreover, as is shown in section 4.4, the amount of prior information has significant influence on the resulting sample size. Smaller sample sizes result from more precise priors whereas diffuse priors lead to larger trials. Therefore, incorporating prior information that properly reflects experts' opinions is the foundation of our Bayesian optimal designs, and strong prior opinion can reduce the size of a study.

In this paper, we do not consider adaptive or sequential design where observations made during the trial can dynamically modify the subsequent trial design. Rather, we provide the Bayesian optimal designs that compute the sample sizes assuming one stage. This is necessary for many investigators because a single sample size is important for practical purposes, such as planning the number of patients to recruit, as well as preparing a budget for a trial. However, early stopping rules can be incorporated into our designs so that a trial can terminate early due to efficacy or futility. One idea is to set the total sample size produced by the proposed Bayesian designs as the maximum number of accrued patients of a phase II study, as suggested by Berger (1985). After obtaining the first m observations, a stopping rule can be evaluated, for example, in terms of the posterior probability that μ_E exeeds μ_C based on the data accumulated so far. In conclusion, the proposed Bayesian optimal designs are very flexible and useful, and can be easily employed when conducting two-arm randomized Phase II clinical trials.

Chapter 5

Summary and Future Directions

The most important studies for developing novel treatments are the clinical trials where statistics play a crucial and indispensable role since the vital feature of clinical research is to generalize statistical inference for the results of new treatments from a limited sample to the general population of patients (Pocock, 2006). This motivated us to concentrate in improving the statistical practice and designs for single-stage two-arm randomized phase II trials.

In this dissertation, we develop an R package called **Sample.Size** based on the optimal designs proposed in Mayo et al. (2010) to accelerate, facilitate, and improve the statistical practice in phase II studies. This package can be used to design oncology phase II trials aiming to estimate response rates in the control and experimental arms and the difference between the two rates with various levels of precision using standard errors as constraints. It is also helpful in designing clinical studies attempting to control precision level of estimates for one proportion only, or the difference between two proportions only, or both proportions but not the difference.

Many probability distributions from the exponential family are important in the practice of phase II clinical trials. Since the original designs developed by Mayo et al. (2010) are limited to trials with dichotomous endpoints only, we develop optimal designs from a frequentist perspective for two-arm randomized phase II clinical trials with endpoints from the exponential dispersion family. The proposed designs are more broadly applicable to other types of commonly used study measures in phase II studies including the normal, exponential, gamma, and Poisson distributions. It is very flexible in that both fixed and flexible allocation ratios are allowed, and can be readily implemented using current software such as SAS and R. Similar to other frequentist methods,

however, the proposed designs are incapable of incorporating the uncertainty inherent in unknown parameters into sample size optimization, and misspecification of these values may lead to a less than optimal design.

Bayesian approaches offer a flexible way to incorporate uncertainty in parameters of interest into considerations. Therefore, we further improve the two frequentist designs to the entire exponential family from a Bayesian perspective. The proposed Bayesian designs use the natural conjugate prior family to optimize the total sample sizes under pre-specified constraints on the expected length of posterior credible intervals for both group means and their difference. It inherits the advantages of the two frequentist approaches for which both fixed and optimal allocation ratio designs are integrated. In addition, the proposed Bayesian designs can be easily implemented when conducting two-arm randomized phase II clinical trials. However, similar to many other Bayesian approaches, the amount of prior information significantly affects the resulting sample sizes, and misleading results may be produced if chosen priors are inappropriate. Therefore, eliciting a prior that properly reflects experts' beliefs is the foundation of the proposed Bayesian designs. Another limitation associated with the proposed approaches is that prior beliefs sometimes cannot be fully described by conjugate priors.

There are several topics that attract our attention for future studies. First, in this dissertation, an R package is only developed for Mayo's optimal designs (Mayo et al., 2010) which are limited to dichotomous outcomes. Both frequentist and Bayesian extensions developed in this dissertation are more broadly applicable to phase II trials with endpoints from the exponential family. It is also necessary to develop R packages for these two proposed designs. Second, when the primary endpoints in phase II trials are time-to-event measures, it is quite often that outcomes of some subjects cannot be observed after a period of time. So it is necessary to extend our current work for phase II clinical trials to designs allowing for incorporating censored observations. Third, for the generalized Bayesian optimal designs, all hyper-parameters are assumed having known and fixed values. The flexibility can be strengthened by extending current Bayesian approaches to designs based on hierarchical models. Fourth, the proposed Bayesian optimal designs are based on the

natural conjugate prior family to which one limitation is that some prior beliefs cannot be fully described by conjugate priors. Therefore, it is beneficial to extend the current work to Bayesian designs without assuming conjugacy. Fifth, both Bayesian and frequentist optimal designs compute the total sample sizes assuming one stage. When primary endpoints in phase II trials are rapidly observable, the sequential or adaptive designs may be more appropriate due to the ethical and economy considerations. So it is beneficial to extend the current work to designs for phase II trials with stopping rules and sample size reestimation.

References

Adcock, C. (1988). A bayesian approach to calculating sample sizes. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37(4/5), 433–439.

Adcock, C. (1997). Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2), 261–283.

Agresti, A. (2014). *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons.

Ameri, A., Poisson, M., Chauveinc, L., Chen, Q., & Delattre, J. (1997). Treatment of recurrent malignant supratentorial gliomas with the association of carboplatin and etoposide: a phase ii study. *Journal of neuro-oncology*, 32(2), 155–160.

Ameri, A., Poisson, M., Chen, Q., & Delattre, J. (1993). Treatment of recurrent malignant supratentorial gliomas with the association of procarbazine, thiotepa and vincristine: a phase ii study. *Journal of neuro-oncology*, 17(1), 43–46.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York, NY: Springer.

Berk, R. & MacDonald, J. M. (2008). Overdispersion and poisson regression. *Journal of Quantitative Criminology*, 24(3), 269–284.

Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2010). *Bayesian adaptive methods for clinical trials*. Boca Raton, FL: CRC press.

Biswas, A., Datta, S., Fine, J. P., & Segal, M. R. (2007). *Statistical advances in the biomedical sciences: clinical trials, epidemiology, survival analysis, and bioinformatics*. Hoboken, NJ: John Wiley & Sons.

Blackwelder, W. C. (1986). Sample size considerations in vaccine trials. *Controlled Clinical Trials*, 7(3), 253.

Brittain, E. & Schlesselman, J. J. (1982). Optimal allocation for the comparison of proportions. *Biometrics*, 38(4), 1003–1009.

Brooks, R. (1987). Optimal allocation for bayesian inference about an odds ratio. *Biometrika*, 74(1), 196–199.

Brunier, H. C. & Whitehead, J. (1994). Sample sizes for phase ii clinical trials derived from bayesian decision theory. *Statistics in Medicine*, 13(23), 2493–2502.

Burtness, B., Powell, M., Catalano, P., Berlin, J., Liles, D. K., Chapman, A. E., Mitchell, E., & Benson, A. B. (2014). Randomized phase ii trial of irinotecan/docetaxel or irinotecan/docetaxel plus cetuximab for metastatic pancreatic cancer: An eastern cooperative oncology group study. *American journal of clinical oncology*. doi: 10.1097/COC.0000000000000068.

Cantor, A. & Moffitt Cancer Center (2009). A sas macro to design phase ii clinical trials. Retrieved from http://analytics.ncsu.edu/sesug/2003/SD09-Cantor.pdf.

Cao, J., Lee, J. J., & Alber, S. (2009). Comparison of bayesian sample size criteria: Acc, alc, and woc. *Journal of Statistical Planning and Inference*, 139(12), 4111–4122.

Casella, G. & Berger, R. L. (2002). *Statistical inference*. Pacific Grove, CA: Duxbury.

Chen, D. & Peace, K. E. (2010). *Clinical Trial Data Analysis Using R*. Boca Raton, FL: CRC press.

Chen, S., Soong, S.-j., & Wheeler, R. H. (1994). An efficient multiple-stage procedure for phase ii clinical trials that have high response rate objectives. *Controlled clinical trials*, 15(4), 277–283.

Chow, S.-C., Chang, M., et al. (2008). Adaptive design methods in clinical trials-a review. *Orphanet Journal of Rare Diseases*, 3(11).

Chung, S. C. & Schulz, M. (2007). Bayesian designs for clinical trials in early drug development. *Journal of Clinical Research Best Proactices*, 3(6), 1–5.

Clarke, B. & Yuan, A. (2006). Closed form expressions for bayesian sample size. *The Annals of Statistics*, 34(3), 1293–1330.

Ding, M., Rosner, G. L., & Müller, P. (2008). Bayesian optimal design for phase ii screening trials. *Biometrics*, 64(3), 886–894.

Eisenhauer, E., Therasse, P., Bogaerts, J., Schwartz, L., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., et al. (2009). New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer*, 45(2), 228–247.

Farewell, V. T., Tom, B. D. M., & Royston, P. (2004). The impact of dichotomization on the efficiency of testing for an interaction effect in exponential family models. *Journal of the American Statistical Association*, 99(467), 822–831.

Farrington, C. P. & Manning, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in medicine*, 9(12), 1447–1454.

Fazal, S. S. (1983). Optimal allocation for the comparison of proportions. *Biometrics*, 39, 1115.

Fleming, T. R., Harrington, D. P., & O'Brien, P. C. (1984). Designs for group sequential tests. *Controlled clinical trials*, 5(4), 348–361.

Food and Drug Administration (2010). Guidance for the use of bayesian statistics in medical device clinical trials. Retrieved from http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf.

Fosgate, G. T. (2009). Practical sample size calculations for surveillance and diagnostic investigations. *Journal of veterinary diagnostic investigation*, 21(1), 3–14.

Friedman, L. M., Furberg, C., & DeMets, D. L. (2010). *Fundamentals of clinical trials*. New York, NY: Springer.

Gadola, L., Poggi, C., Poggio, M., Sáez, L., Ferrari, A., Romero, J., Fumero, S., Ghelfi, G., Chiffflet, L., & Borges, P. L. (2013). Using a multidisciplinary training program to reduce peritonitis in peritoneal dialysis patients. *Peritoneal Dialysis International*, 33(1), 38–45.

Gajewski, B. J. & Mayo, M. S. (2006). Bayesian sample size calculations in phase ii clinical trials using a mixture of informative priors. *Statistics in medicine*, 25(15), 2554–2566.

Gehan, E. A. (1961). The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *Journal of chronic diseases*, 13(4), 346–353.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. Boca Raton, FL: CRC press.

Groulx, A., Moon, K.-h., & Chung, S. (2007). Using sas to determine sample sizes for traditional 2-stage and adaptive 2-stage phase ii clinical trial designs. Retrieved from http://www2.sas.com/proceedings/forum2007/188-2007.pdf.

Hand, A. L., Stamey, J. D., & Young, D. M. (2011). Bayesian sample-size determination for two independent poisson rates. *Computer methods and programs in biomedicine*, 104(2), 271–277.

Herndon, J. E. (1998). A design alternative for two-stage, phase ii, multicenter cancer clinical trials. *Controlled clinical trials*, 19(5), 440–450.

Jiang, W., Mahnken, J. D., He, J., & Mayo, M. S. (2014). Generalized optimal designs for two-arm, randomized phase ii clinical trials with endpoints from the exponential dispersion families. Manuscript submitted to Statistics in Medicine for publication.

Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T., Grosbein, H. A., & Feldman, B. M. (2010). A valid and reliable belief elicitation method for bayesian priors. *Journal of clinical epidemiology*, 63(4), 370–383.

Joseph, L. & Belisle, P. (1997). Bayesian sample size determination for normal means and differences between normal means. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2), 209–226.

Joseph, L., Du Berger, R., & Belisle, P. (1997). Bayesian and mixed bayesian/likelihood criteria for sample size determination. *Statistics in Medicine*, 16(7), 769–781.

Joseph, L. & Wolfson, D. B. (1997). Interval-based versus decision theoretic criteria for the choice of sample size. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2), 145–149.

Joseph, L., Wolfson, D. B., & Du Berger, R. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 44(2), 143–154.

Jung, S.-H. (2008). Randomized phase ii trials with a prospective control. *Statistics in medicine*, 27(4), 568–583.

Jung, S.-H. (2013). *Randomized phase II cancer clinical trials*. Boca Raton: CRC press.

Karrison, T. G., Maitland, M. L., Stadler, W. M., & Ratain, M. J. (2007). Design of phase ii cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non–small-cell lung cancer. *Journal of the National Cancer Institute*, 99(19), 1455–1461.

Lavin, P. T. (1981). An alternative model for the evaluation of antitumor activity. *Cancer clinical trials*, 4(4), 451–457.

Lin, Y. & Shih, W. J. (2004). Adaptive two-stage designs for single-arm phase iia cancer clinical trials. *Biometrics*, 60(2), 482–490.

Lindley, D. V. (1997). The choice of sample size. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2), 129–138.

Mahajan, R. & Gupta, K. (2010). Adaptive design clinical trials: Methodology, challenges and prospect. *Indian journal of pharmacology*, 42(4), 201.

Mandrekar, S. J. & Sargent, D. J. (2010). Randomized phase ii trials: time for a new era in clinical trial design. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer*, 5(7), 932–934.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537–563.

Mayo, M. S. & Gajewski, B. J. (2004). Bayesian sample size calculations in phase ii clinical trials using informative conjugate priors. *Controlled clinical trials*, 25(2), 157–167.

Mayo, M. S., Mahnken, J. D., & Soong, S.-j. (2010). Optimal designs for two-arm, phase ii clinical trial design with multiple constraints. *Journal of biopharmaceutical statistics*, 20(1), 106–124.

M'Lan, C. E., Joseph, L., Wolfson, D. B., et al. (2008). Bayesian sample size determination for binomial proportions. *Bayesian Analysis*, 3(2), 269–296.

O'Brien, M., Montes, A., & Powles, T. (1996). Hormone replacement therapy as treatment of breast cancer–a phase ii study of org od 14 (tibilone). *British journal of cancer*, 73(9), 1086–1088.

O'Brien, P. C. & Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, 35(3), 549–556.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*. Chichester, England: John Wiley & Sons.

Pezeshk, H. (2002). How many subjects?–a bayesian approach to the design of a clinical trial. *Iranian International Journal of Science*, 3(1), 127–133.

Pham-Gia, T. (1997). On bayesian analysis, bayesian decision theory and the sample size problem. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(2), 139–144.

Pham-Gia, T. & Turkkan, N. (2003). Determination of exact sample sizes in the bayesian estimation of the difference of two proportions. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(2), 131–150.

Piantadosi, S. (2005). *Clinical trials: a methodologic perspective*. Hoboken, NJ: John Wiley & Sons.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2), 191–199.

Pocock, S. J. (2006). *Clinical trials: a practical approach*. Chichester, England: John Wiley & Sons.

Raiffa, H. & Schlaifer, R. (1961). *Applied statistical decision theory*. Boston, MA: Clinton Press.

Rubinstein, L., Crowley, J., Ivy, P., LeBlanc, M., & Sargent, D. (2009). Randomized phase ii designs. *Clinical Cancer Research*, 15(6), 1883–1890.

Rubinstein, L. V., Korn, E. L., Freidlin, B., Hunsberger, S., Ivy, S. P., & Smith, M. A. (2005). Design issues of randomized phase ii trials and a proposal for phase ii screening trials. *Journal of Clinical Oncology*, 23(28), 7199–7206.

Sahai, H. & Khurshid, A. (1996). Formulae and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: A review. *Statistics in Medicine*, 15(1), 1–21.

Sahu, S. & Smith, T. (2006). A bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(2), 235–253.

Sanson, M., Ameri, A., Monjour, A., Sahmoud, T., Ronchin, P., Poisson, M., & Delattre, J. (1996). Treatment of recurrent malignant supratentorial gliomas with ifosfamide, carboplatin and etoposide: a phase ii study. *European Journal of Cancer*, 32(13), 2229–2235.

Sargent, D. J., Sloan, J. A., & Cha, S. S. (1999). Sample size and design considerations for phase ii clinical trials with correlated observations. *Controlled clinical trials*, 20(3), 242–252.

Simon, R. (1989). Optimal two-stage designs for phase ii clinical trials. *Controlled clinical trials*, 10(1), 1–10.

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. Chichester, England: John Wiley & Sons.

Stallard, N. (1998). Sample size determination for phase ii clinical trials based on bayesian decision theory. *Biometrics*, 54(1), 279–294.

Stallard, N. & Thall, P. F. (2001). Decision-theoretic designs for pre-phase ii screening trials in oncology. *Biometrics*, 57(4), 1089–1095.

Stallard, N., Thall, P. F., & Whitehead, J. (1999). Decision theoretic designs for phase ii clinical trials with multiple outcomes. *Biometrics*, 55(3), 971–977.

Staquet, M. & Sylvester, R. (1977). A decision theory approach to phase ii clinical trials. *Biomedicine*, 26(4), 262–266.

Sun, L. Z., Chen, C., & Patel, K. (2009). Optimal two-stage randomized multinomial designs for phase ii oncology trials. *Journal of biopharmaceutical statistics*, 19(3), 485–493.

Sylvester, R. J. (1988). A bayesian approach to the design of phase ii clinical trials. *Biometrics*, 44(3), 823–836.

Tan, S.-B. & Machin, D. (2002). Bayesian two-stage designs for phase ii clinical trials. *Statistics in medicine*, 21(14), 1991–2012.

Taylor, J. M., Braun, T. M., & Li, Z. (2006). Comparing an experimental agent to a standard agent: relative merits of a one-arm or randomized two-arm phase ii design. *Clinical Trials*, 3(4), 335–348.

Thall, P. F. (2008). A review of phase 2–3 clinical trial designs. *Lifetime data analysis*, 14(1), 37–53.

Thall, P. F. & Cheng, S.-C. (1999). Treatment comparisons based on two-dimensional safety and efficacy alternatives in oncology trials. *Biometrics*, 55(3), 746–753.

Thall, P. F. & Cheng, S.-C. (2001). Optimal two-stage designs for clinical trials based on safety and efficacy. *Statistics in medicine*, 20(7), 1023–1032.

Thall, P. F. & Simon, R. (1994). Practical bayesian guidelines for phase iib clinical trials. *Biometrics*, 50(2), 337–349.

Thall, P. F., Simon, R. M., & Estey, E. H. (1995). Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statistics in medicine*, 14(4), 357–379.

Verhey, L. H., Signori, A., Arnold, D. L., Bar-Or, A., Sadovnick, A. D., Marrie, R. A., Banwell, B., & Sormani, M. P. (2013). Clinical and mri activity as determinants of sample size for pediatric multiple sclerosis trials. *Neurology*, 81(14), 1215–1221.

Walter, S. D. (1977). Determination of significant relative risks and optimal sampling procedures in prospective and retrospective comparative studies of various sizes. *American journal of epidemiology*, 105(4), 387–397.

Wason, J., Mander, A. P., & Thompson, S. G. (2012). Optimal multistage designs for randomised clinical trials with continuous outcomes. *Statistics in medicine*, 31(4), 301–312.

Wason, J. M. & Mander, A. P. (2012). Minimizing the maximum expected sample size in two-stage phase ii clinical trials with continuous outcomes. *Journal of biopharmaceutical statistics*, 22(4), 836–852.

Whitehead, J., Valdés-Márquez, E., & Lissmats, A. (2009). A simple two-stage design for quantitative responses with application to a study in diabetic neuropathic pain. *Pharmaceutical statistics*, 8(2), 125–135.

Wieand, H. S. (2005). Randomized phase ii trials: what does randomization gain? *Journal of clinical oncology*, 23(9), 1794–1795.

Zhang, X., Cutter, G., & Belin, T. (2011). Bayesian sample size determination under hypothesis tests. *Contemporary clinical trials*, 32(3), 393–398.

Zhong, W., Koopmeiners, J. S., & Carlin, B. P. (2013). A two-stage bayesian design with sample size reestimation and subgroup analysis for phase ii binary response trials. *Contemporary clinical trials*, 36(2), 587–596.

Zohar, S., Teramukai, S., & Zhou, Y. (2008). Bayesian design and conduct of phase ii single-arm clinical trials with binary outcomes: a tutorial. *Contemporary clinical trials*, 29(4), 608–616.